GRADIENT BOOST

Селезнёв Артём
Аналитик больших данных
2019

МегаФон

Градиентный бустинг скрывает в себе

1. $Set(x_i, x_j)_{i=1}^{n}$ и функцию потерь $L(y_i, F(x))$
2. $F_0(x) = argmin\Sigma L(y_i, \gamma)$
3. Вычисление $r_{i,m} = -[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}]_{F(x)=F_{m-1}(x)}$
4. Обучение регрессии r_{jm} в цикле $j = 1 \ldots m$
5. В цикле $j = 1 \ldots m$ вычисление $\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$
6. $F_m(x) = F_{m-1}(x) + \nu\Sigma\gamma_m I(x \in R_{jm})$

1

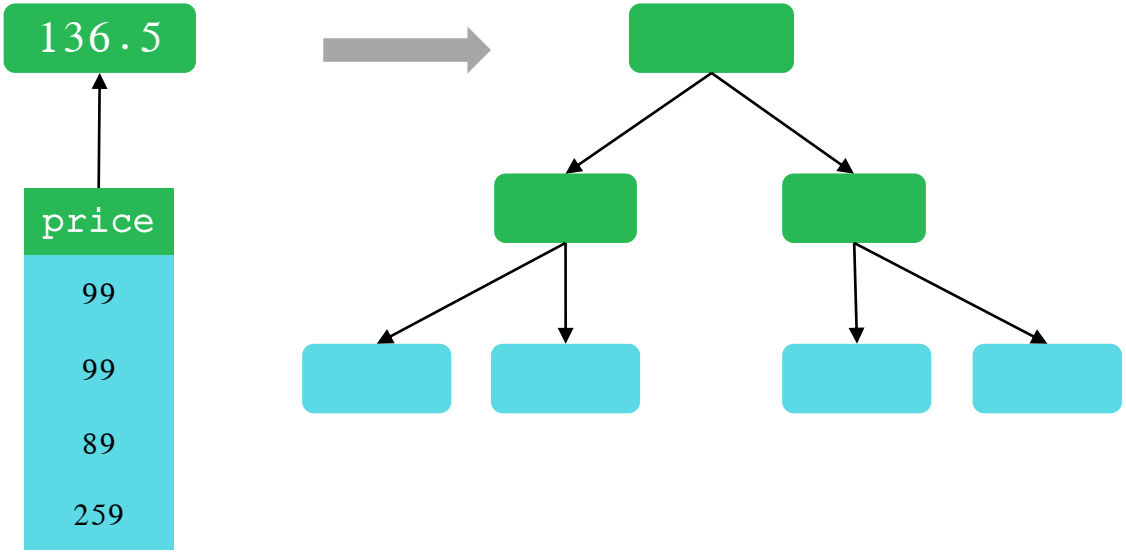
Разберем
на примере.
Что скрыто
в ARGMIN?

Наш помощник – набор данных

Region	Category	Subcategory	Param1	threshold	price
Владимирская область	Личные вещи	1	1	3	99
Волгоградская область	Личные вещи	2	1	2	99
Кировская область	Личные вещи	5	1	3	89
Москва	Личные вещи	4	6	3	259

Вспомните деревья

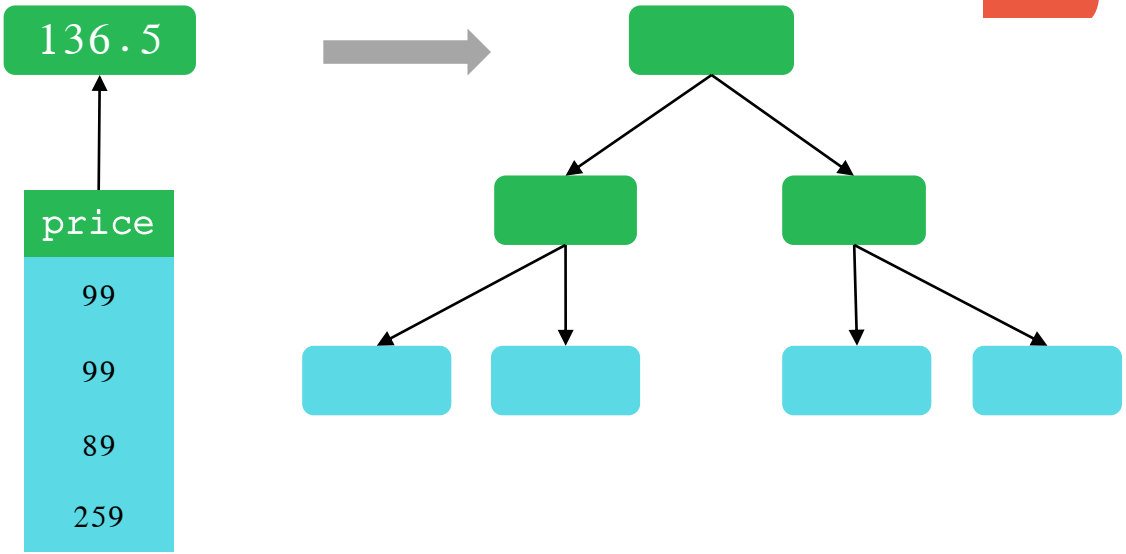
Вспомните деревья



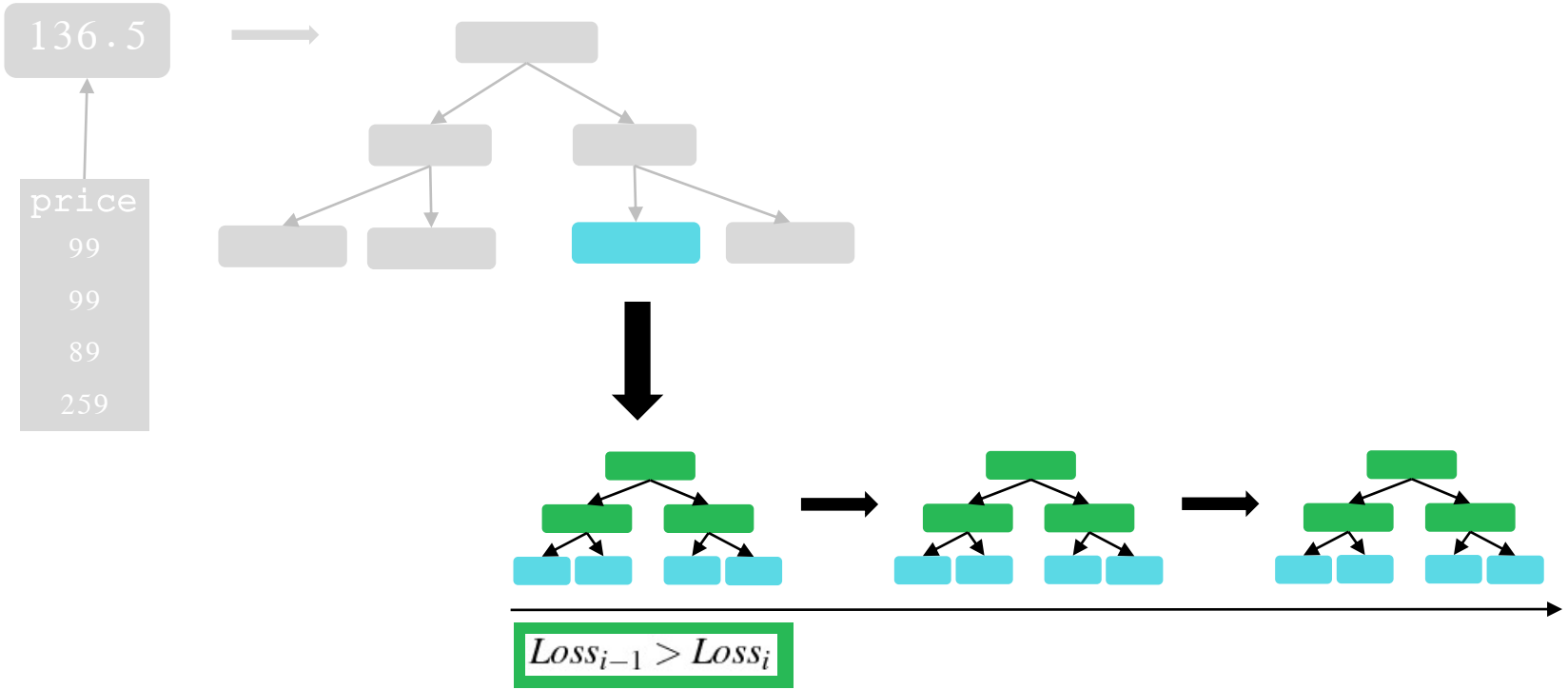
Вспомните деревья

Глубина должна быть ограничена.

Обычно 8 – 32



Вспомните деревья



Начинаем считать

price
99
99
89
259

→ 136,5

(Observed − Predicted)
= (99 − 136.5) = −37.5

Region	Category	Sub category	Param1	threshold	price	diff
Владимирская область	Личные вещи	1	1	3	99	
Волгоградская область	Личные вещи	2	1	2	99	
Кировская область	Личные вещи	5	1	3	89	
Москва	Личные вещи	4	6	3	259	

Начинаем считать

price
99
99
89
259

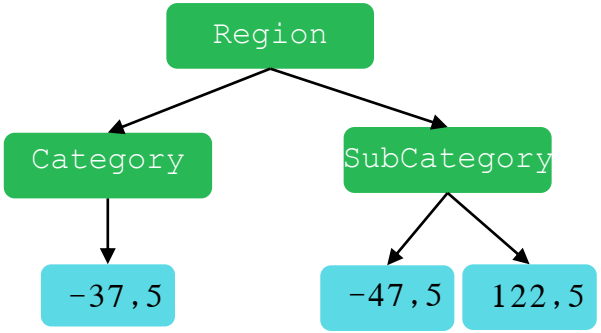
→ 136,5

(Observed − Predicted)
= (99 − 136.5) = −37.5

Region	Category	Sub category	Param1	threshold	price	diff
Владимирская область	Личные вещи	1	1	3	99	−37,5
Волгоградская область	Личные вещи	2	1	2	99	−37,5
Кировская область	Личные вещи	5	1	3	89	−47,5
Москва	Личные вещи	4	6	3	259	122,5

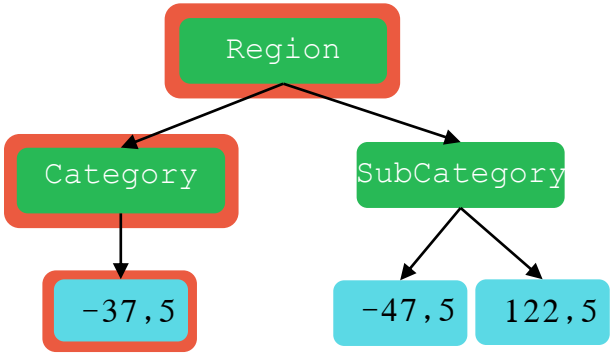
Делаем первое дерево

Region	Category	Subcategory	Param1	threshold	price	T
Владимирская область	Личные вещи	1	1	3	99	-37,5
Волгоградская область	Личные вещи	2	1	2	99	-37,5
Кировская область	Личные вещи	5	1	3	89	-47,5
Москва	Личные вещи	4	6	3	259	122,5



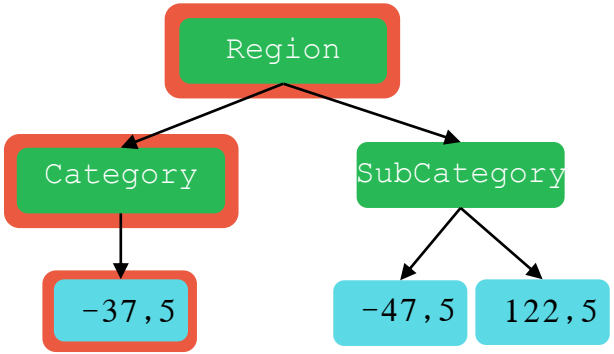
Делаем первое дерево

Region	Category	Subcategory	Param1	threshold	price	T
Владимирская область	Личные вещи	1	1	3	99	-37,5
Волгоградская область	Личные вещи	2	1	2	99	-37,5
Кировская область	Личные вещи	5	1	3	89	-47,5
Москва	Личные вещи	4	6	3	259	122,5



Делаем первое дерево

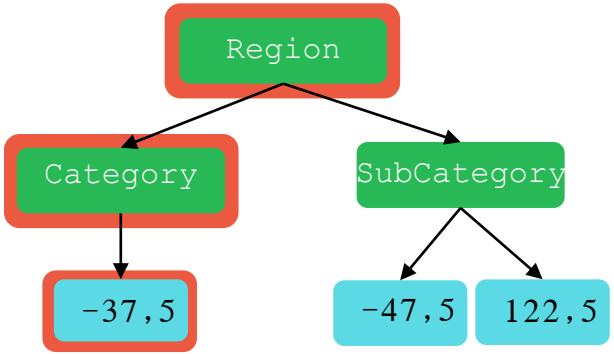
Region	Category	Subcategory	Param1	threshold	price	T
Владимирская область	Личные вещи	1	1	3	99	-37,5
Волгоградская область	Личные вещи	2	1	2	99	-37,5
Кировская область	Личные вещи	5	1	3	89	-47,5
Москва	Личные вещи	4	6	3	259	122,5



$$(\text{Avg Price} + \text{Predicted}) = (136.5 + (-37.5)) = 99$$

Делаем первое дерево

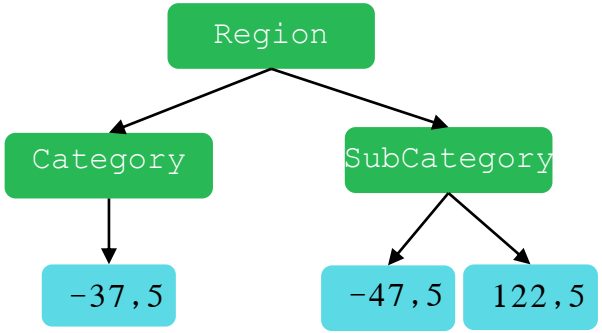
Region	Category	Subcategory	Param1	threshold	price	T
Владимирская область	Личные вещи	1	1	3	99	-37,5
Волгоградская область	Личные вещи	2	1	2	99	-37,5
Кировская область	Личные вещи	5	1	3	89	-47,5
Москва	Личные вещи	4	6	3	259	122,5



(Avg Price + Predicted)
= (136.5 + (- 37.5)) =

99

Делаем первое дерево

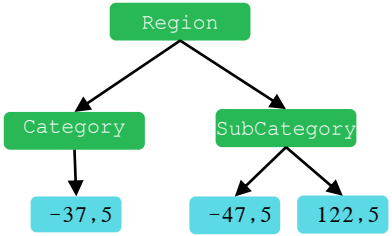


2

Надо использовать коэффициент скорости обучения

```
(Avg Price + (Learning Rate * Predicted))
= 136.5 + (0.1 * (- 37.5)) = 132.75
```

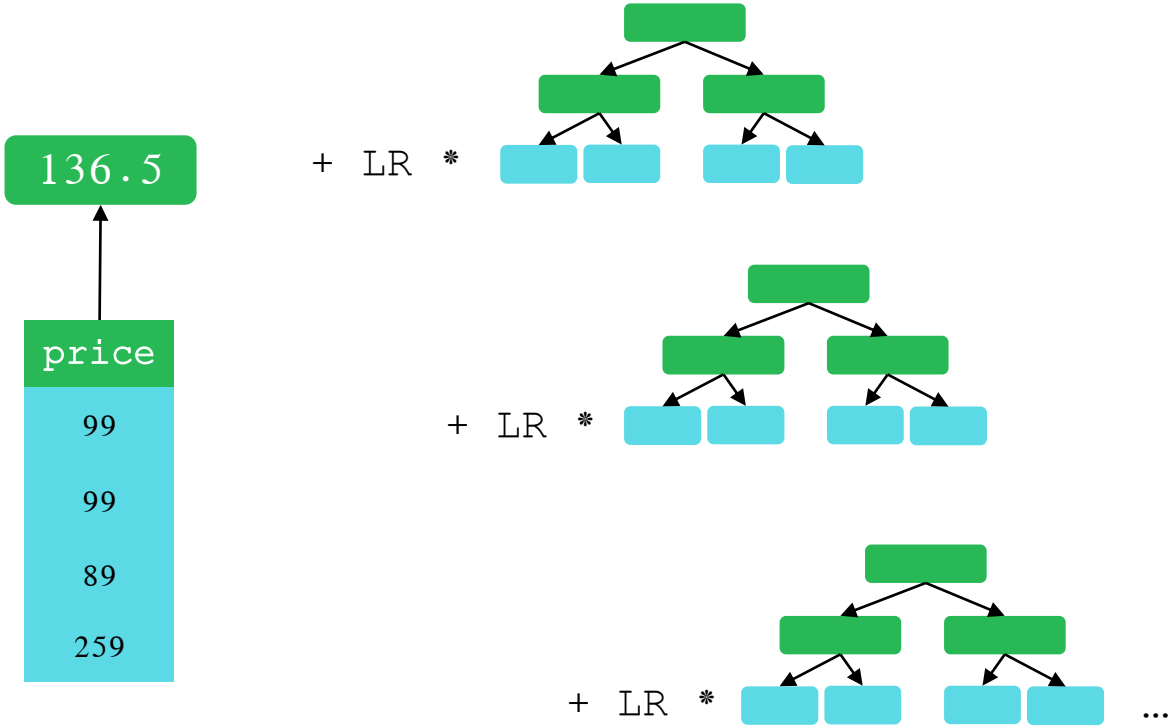
Делаем первое дерево



(Avg Price +
(Learning Rate * Predicted))

Region	Category	Subcategory	Param1	threshold	price	diff
Владимирская область	Личные вещи	1	1	3	99	-37,5
Волгоградская область	Личные вещи	2	1	2	99	-37,5
Кировская область	Личные вещи	5	1	3	89	-47,5
Москва	Личные вещи	4	6	3	259	122,5

Ещё больше деревьев



*LR – learning rate

Смотрите

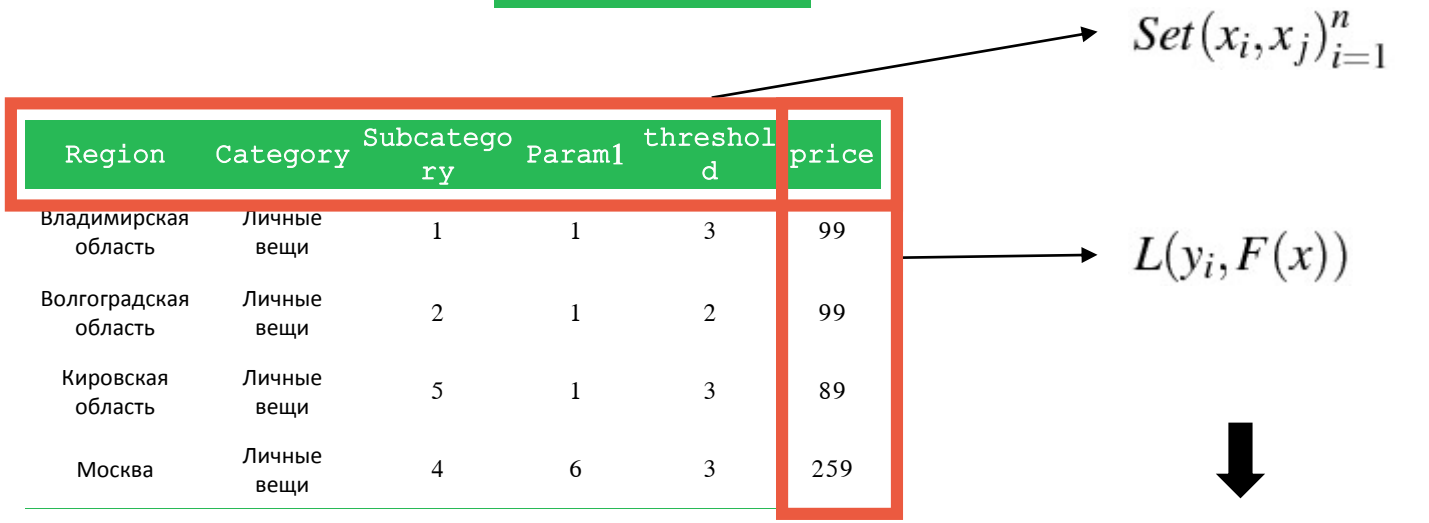
Мы разобрали:

1. $Set(x_i, x_j)_{i=1}^{n}$ и функцию потерь $L(y_i, F(x))$
2. $F_0(x) = argmin \Sigma L(y_i, \gamma)$
3. Вычисление $r_{i,m} = -[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}]_{F(x)=F_{m-1}(x)}$
4. Обучение регрессии r_{jm} в цикле $j = 1 \ldots m$
5. В цикле $j = 1 \ldots m$ вычисление $\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$
6. $F_m(x) = F_{m-1}(x) + \nu \Sigma \gamma_m I(x \in R_{jm})$

Давайте закрепим

Наш помощник – набор данных

$$Set(x_i, x_j)_{i=1}^{n}$$



$$Set(x_i, x_j)_{i=1}^{n}$$

Region	Category	Subcategory	Param1	threshold	price
Владимирская область	Личные вещи	1	1	3	99
Волгоградская область	Личные вещи	2	1	2	99
Кировская область	Личные вещи	5	1	3	89
Москва	Личные вещи	4	6	3	259

$$L(y_i, F(x))$$

(Observed – Predicted)

Наш помощник – набор данных

$$Set(x_i, x_j)_{i=1}^{n}$$

Region	Category	Subcategory	Param1	threshold	price
Владимирская область	Личные вещи	1	1	3	99
Волгоградская область	Личные вещи	2	1	2	99
Кировская область	Личные вещи	5	1	3	89
Москва	Личные вещи	4	6	3	259

$$Set(x_i, x_j)_{i=1}^{n}$$

$$L(y_i, F(x))$$

$$\frac{diff}{diff * Predicted} * \frac{1}{2}(Observed - Predicted)^2 = \frac{1}{2}(Observed - Predicted)^2$$

Наш помощник – набор данных

$$Set(x_i, x_j)_{i=1}^{n}$$

Region	Category	Subcategory	Param1	threshold	price
Владимирская область	Личные вещи	1	1	3	99
Волгоградская область	Личные вещи	2	1	2	99
Кировская область	Личные вещи	5	1	3	89
Москва	Личные вещи	4	6	3	259

$$Set(x_i, x_j)_{i=1}^{n}$$

$$L(y_i, F(x))$$

$$\frac{diff}{diff * Predicted} * \frac{1}{2}(Observed - Predicted)^2 = \frac{1}{2}(Observed - Predicted)^2$$

Наш помощник – набор данных

$$Set(x_i, x_j)_{i=1}^{n}$$

Region	Category	Subcategory	Param1	threshold	price
Владимирская область	Личные вещи	1	1	3	99
Волгоградская область	Личные вещи	2	1	2	99
Кировская область	Личные вещи	5	1	3	89
Москва	Личные вещи	4	6	3	259

$$Set(x_i, x_j)_{i=1}^{n}$$

$$L(y_i, F(x))$$

$$\frac{1}{2}(Observed - Predicted)^2 = \frac{2}{2}(Observed - Predicted)x - 1 = -(Observed - Predicted)$$

Наш помощник – набор данных

$$Set(x_i, x_j)_{i=1}^{n}$$

$$Set(x_i, x_j)_{i=1}^{n}$$

Region	Category	Subcategory	Param1	threshold	price
Владимирская область	Личные вещи	1	1	3	99
Волгоградская область	Личные вещи	2	1	2	99
Кировская область	Личные вещи	5	1	3	89
Москва	Личные вещи	4	6	3	259

$$L(y_i, F(x))$$

$$\frac{1}{2}(Observed - Predicted)^2 = \frac{2}{2}(Observed - Predicted)x - 1 = -(Observed - Predicted)$$

Argmin <- минимизируем потерю

$$F_0(x) = argmin \Sigma L(y_i, \gamma)$$

Argmin <- минимизируем потерю

$$F_0(x) = argmin \Sigma L(y_i, \gamma)$$

price
99
99
89
259

⬇

$$\frac{1}{2}(99 - Predicted)^2 + \frac{1}{2}(99 - Predicted)^2 +$$

$$\frac{1}{2}(89 - Predicted)^2 + \frac{1}{2}(259 - Predicted)^2$$

Argmin <- минимизируем потерю

$$F_0(x) = argmin \Sigma L(y_i, \gamma)$$

price
99
99
89
259

$$\frac{1}{2}(99 - Predicted)^2 \rightarrow -(99 - Predicted) +$$

$$\frac{1}{2}(99 - Predicted)^2 \rightarrow -(99 - Predicted) +$$

$$\frac{1}{2}(89 - Predicted)^2 \rightarrow -(89 - Predicted) +$$

$$\frac{1}{2}(259 - Predicted)^2 \rightarrow -(259 - Predicted)$$

Argmin <- минимизируем потерю

$$F_0(x) = argmin \Sigma L(y_i, \gamma)$$

price
99
99
89
259

⬇

$$-99 + Predicted + (-99) + Predicted + (-89) + Predicted + (-259) + Predicted = 0$$

Argmin <- минимизируем потерю

$$F_0(x) = argmin \Sigma L(y_i, \gamma)$$

price
99
99
89
259

⬇

$$-99 + Predicted + (-99) + Predicted + (-89) + Predicted + (-259) + Predicted = 0$$

⬇

$$Predicted + Predicted + Predicted = 99 + 89 + 259$$

Argmin <- минимизируем потерю

$$F_0(x) = argmin \Sigma L(y_i, \gamma)$$

price
99
99
89
259

⬇

$$-99 + Predicted + (-99) + Predicted + (-89) + Predicted + (-259) + Predicted = 0$$

⬇

$$4 * Predicted = (2*99) + 89 + 259$$

⬇

$$\frac{(2*99)+89+259}{4}$$

2

Градиентный спуск

Теперь разберём :

1. $Set(x_i, x_j)_{i=1}^{n}$ и функцию потерь $L(y_i, F(x))$
2. $F_0(x) = argmin\Sigma L(y_i, \gamma)$
3. Вычисление $r_{i,m} = -[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}]_{F(x)=F_{m-1}(x)}$
4. Обучение регрессии r_{jm} в цикле j = 1...m
5. В цикле j = 1...m вычисление $\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$
6. $F_m(x) = F_{m-1}(x) + \nu\Sigma\gamma_m I(x \in R_{jm})$

Сначала посмотрим

Разбираем шаги спуска

$$r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

Разбираем шаги спуска

$$r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

$$\frac{diff}{diff * Predicted} * \frac{1}{2}(Observed - Predicted)^2$$

Разбираем шаги спуска

$$r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

Разбираем шаги спуска

$$r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

$$\downarrow$$

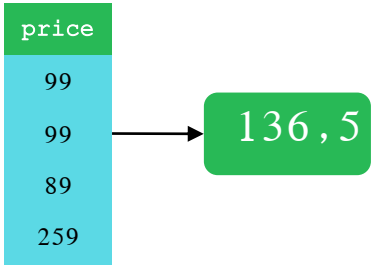
$$\left(\text{Observed } F_{m-1}(x)\right.$$

$$\downarrow$$

$$\left(\text{Observed } F_0(x) \;\middle|\; F_0(x)\right) \rightarrow$$

136.

Помните этот слайд?



(Observed − Predicted)
= (99 − 136.5) = −37.5

Region	Category	Sub category	Param1	threshold	price	diff
Владимирская область	Личные вещи	1	1	3	99	-37,5
Волгоградская область	Личные вещи	2	1	2	99	-37,5
Кировская область	Личные вещи	5	1	3	89	-47,5
Москва	Личные вещи	4	6	3	259	122,5

Он считает R

price
99
99
89
259

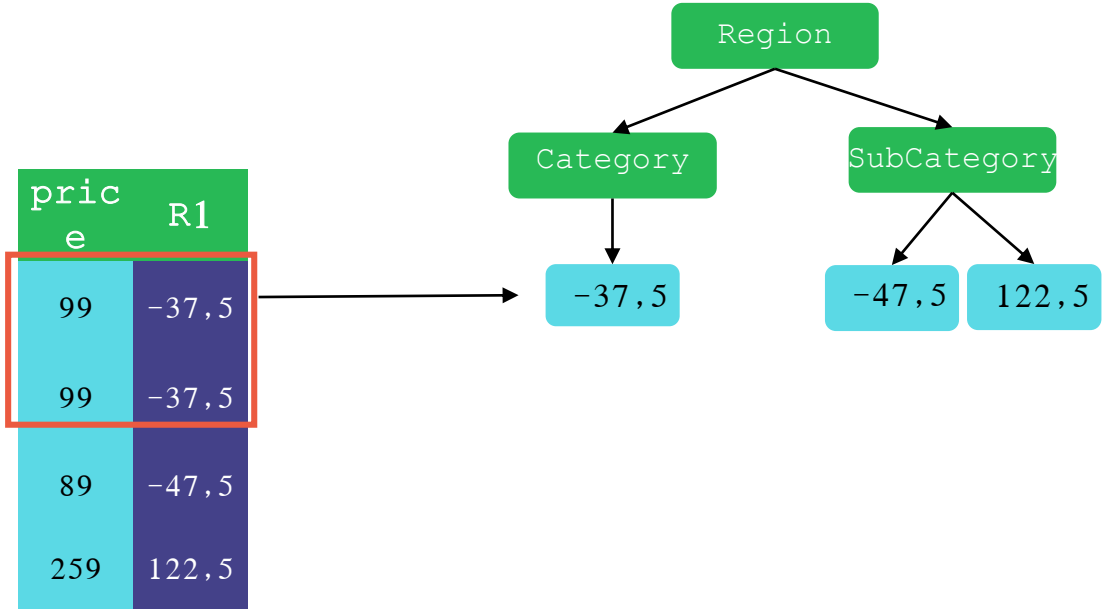
136,5

(Observed − Predicted)
= (99 − 136.5) = −37.5

Region	Category	Sub category	Param1	threshold	price	R1
Владимирская область	Личные вещи	1	1	3	99	−37,5
Волгоградская область	Личные вещи	2	1	2	99	−37,5
Кировская область	Личные вещи	5	1	3	89	−47,5
Москва	Личные вещи	4	6	3	259	122,5

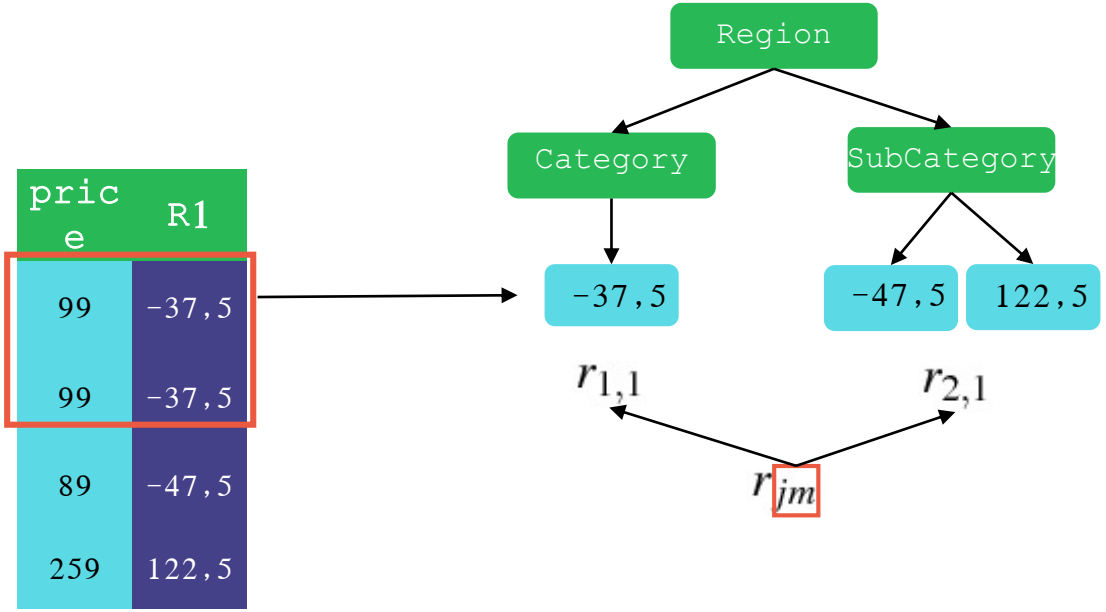
Обучение регрессии r_{jm}

в цикле j = 1..m



Обучение регрессии r_{jm}

в цикле `j = 1..m`



В цикле `j = 1...m`
вычисление `gamma argmin`

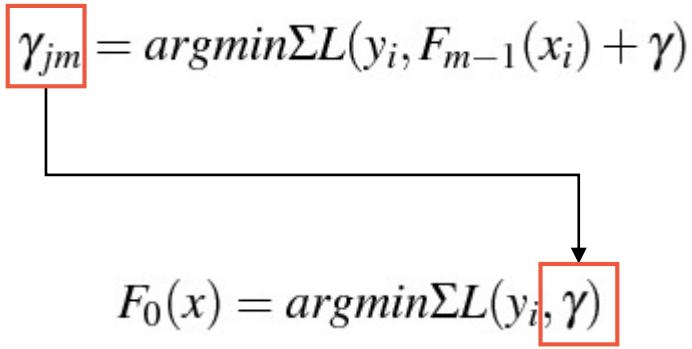
$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

В цикле `j = 1...m`
вычисление `gamma argmin`

$$\gamma_{jm} = \boxed{argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)}$$

$$F_0(x) = \boxed{argmin \Sigma L(y_i, \gamma)}$$

В цикле `j = 1...m`
вычисление `gamma argmin`



В цикле `j = 1...m`
вычисление `gamma argmin`

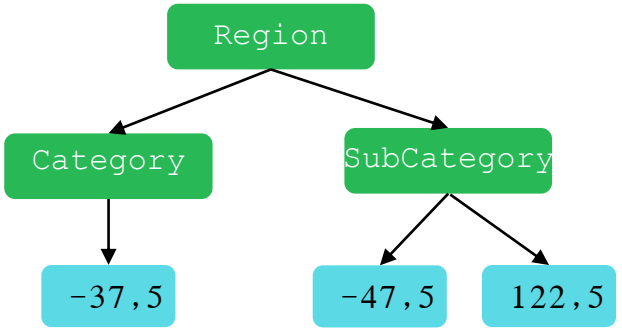
$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

В цикле `j = 1...m`
вычисление `gamma argmin`

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

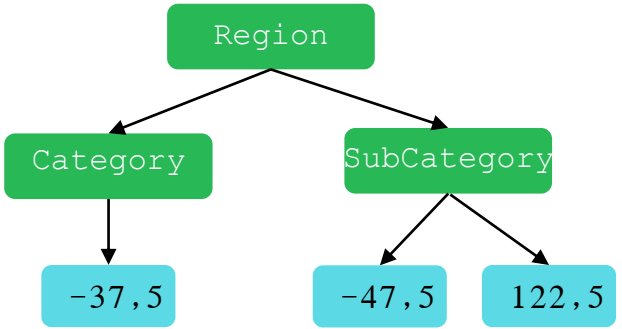
r_{jm}

$r_{1,1}$ $r_{2,1}$



В цикле `j = 1...m`
вычисление `gamma argmin`

$$\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



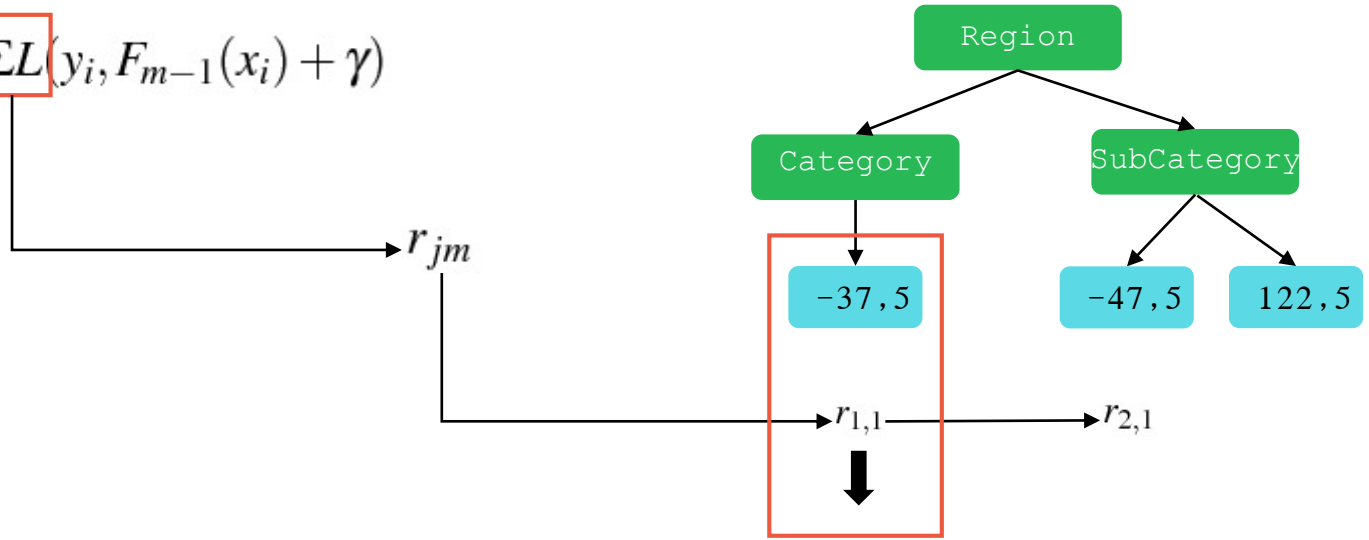
r_{jm}

$r_{1,1}$ $r_{2,1}$

$$\gamma_{1,1} = argmin\frac{1}{2}(y_i - (F_{m-1}(x_i) + \gamma))^2 = argmin\frac{1}{2}(-37.5 - \gamma)^2$$

В цикле `j = 1...m`
вычисление `gamma argmin`

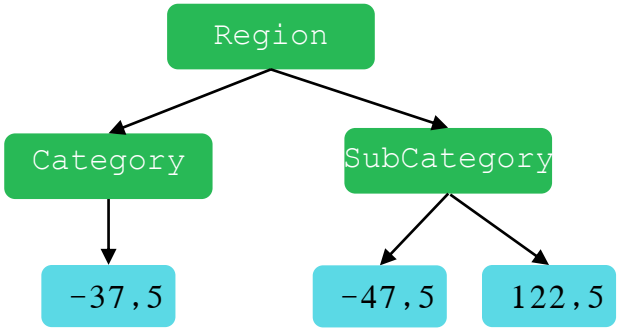
$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{1,1} = argmin \frac{1}{2}(y_i - (F_{m-1}(x_i) + \gamma))^2 = argmin \frac{1}{2}(-37.5 - \gamma)^2$$

В цикле `j = 1...m`
вычисление `gamma argmin`

$$\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



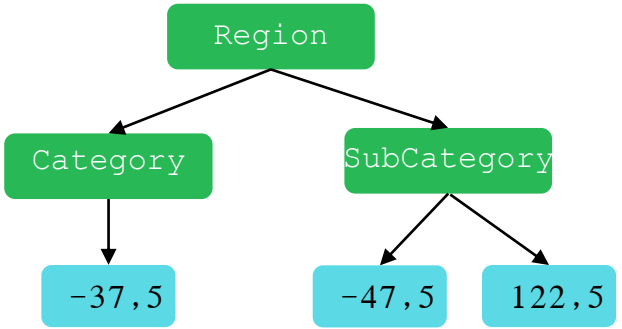
r_{jm}

$r_{1,1}$ $r_{2,1}$

$$\gamma_{1,1} = argmin\frac{1}{2}(y_i - (F_{m-1}(x_i) + \gamma))^2 = argmin\frac{1}{2}(-37.5 - \gamma)^2$$

В цикле `j = 1...m`
вычисление `gamma argmin`

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



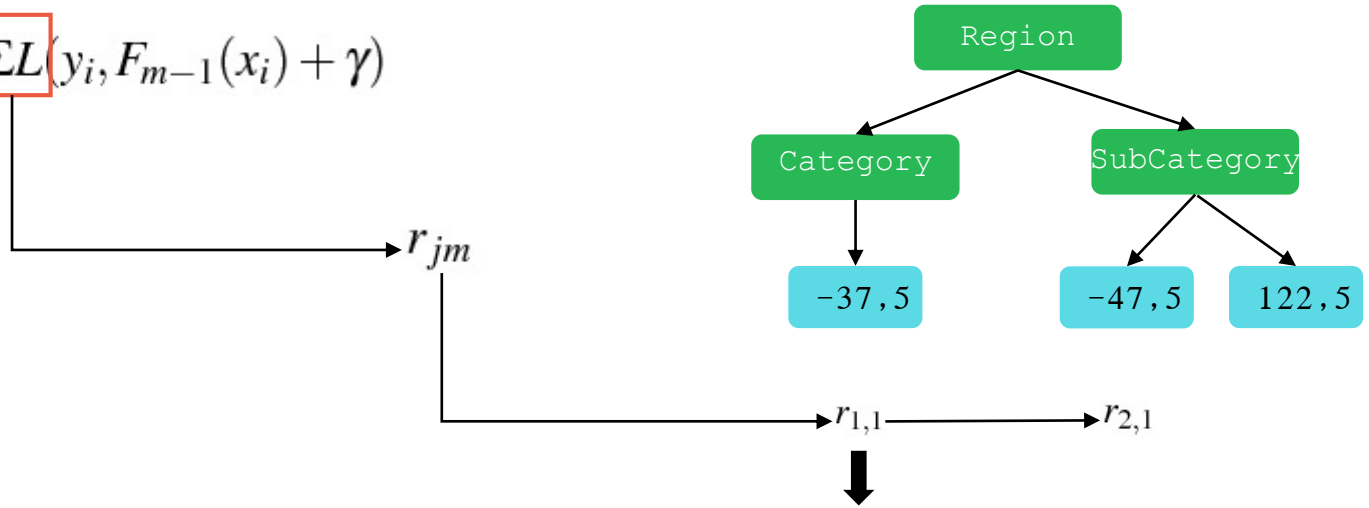
$$r_{jm}$$

$$r_{1,1} \quad r_{2,1}$$

$$\gamma_{1,1} = argmin \frac{1}{2}(y_i - (F_{m-1}(x_i) + \gamma))^2 = argmin \frac{1}{2}(-37.5 - \gamma)^2 \Rightarrow \frac{\partial}{\partial \gamma} \frac{1}{2}(-37.5 - \gamma)^2$$

В цикле `j = 1...m`
вычисление `gamma argmin`

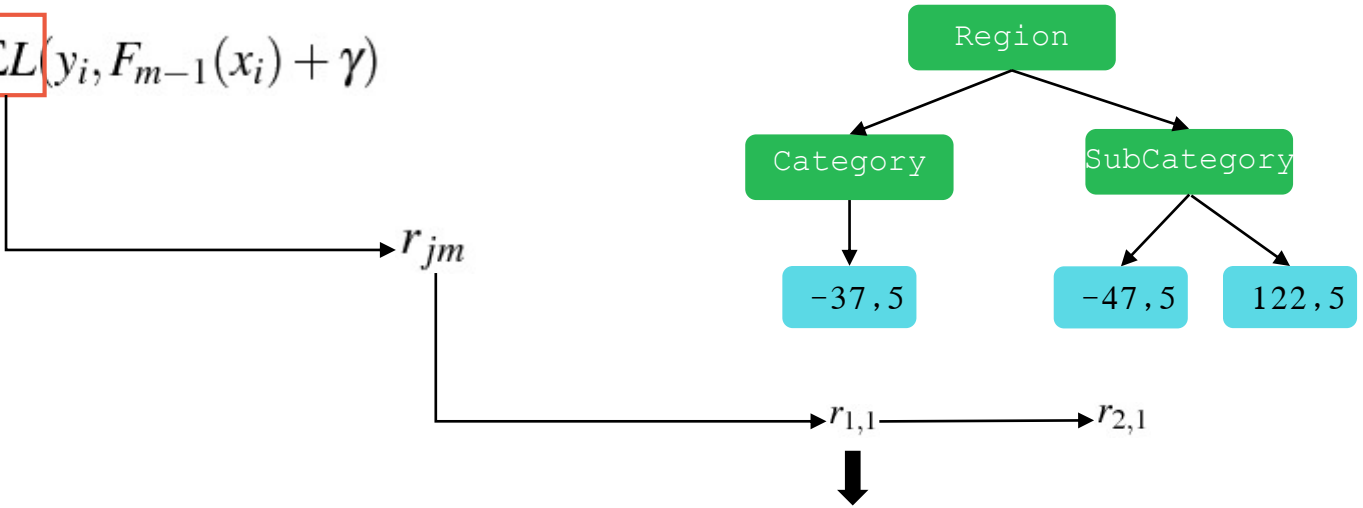
$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\frac{\partial}{\partial \gamma} \frac{1}{2}(-37.5 - \gamma)^2 \rightarrow 37.5 + \gamma = 0$$

В цикле `j = 1...m`
вычисление `gamma argmin`

$$\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\frac{\partial}{\partial\gamma}\frac{1}{2}(-37.5-\gamma)^2 \Rightarrow 37.5+\gamma=0 \Rightarrow -37.5$$

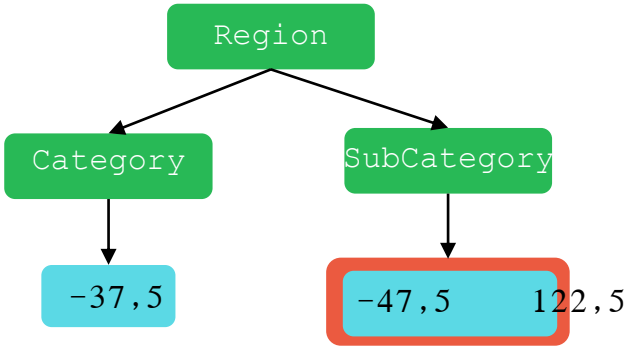
В цикле j = 1…m
вычисление gamma argmin

(если несколько значений в листе)

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

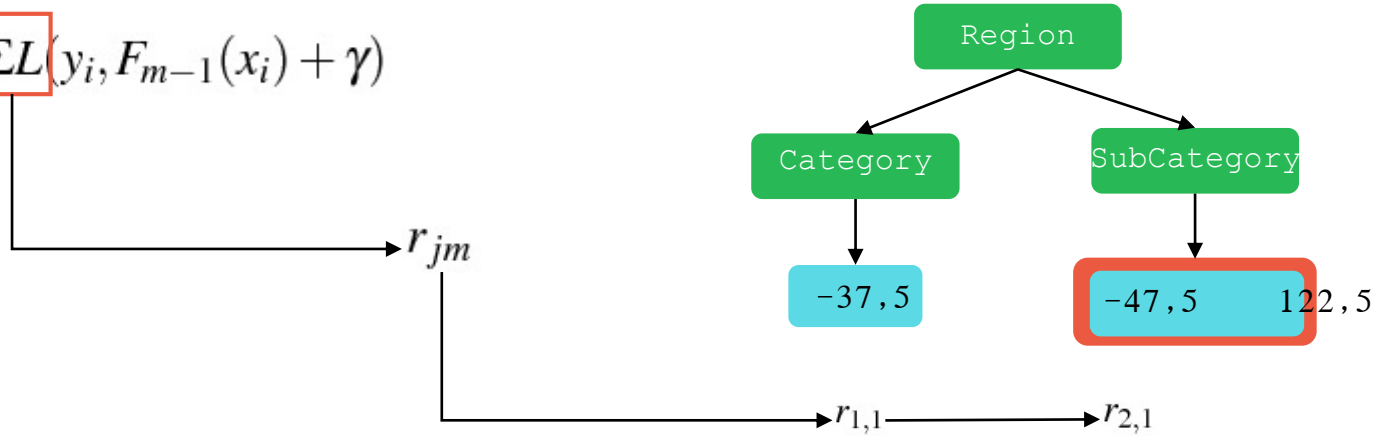
r_{jm}

$r_{1,1}$ $r_{2,1}$



В цикле j = 1…m
вычисление gamma argmin
(если несколько значений в листе)

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

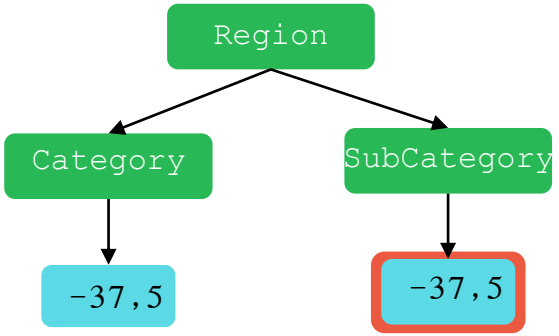


$$47.5 + \gamma - 122.5 + \gamma = 0 \quad \Rightarrow \quad \frac{47.5 - 122.5}{2} = -37.5$$

В цикле j = 1…m
вычисление gamma argmin

(если несколько значений в листе)

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



r_{jm}

$r_{1,1}$ $r_{2,1}$

$$47.5 + \gamma - 122.5 + \gamma = 0 \rightarrow \frac{47.5 - 122.5}{2} = -37.5$$

Обновляем результат

$$F_m(x) = F_{m-1}(x) + \nu\Sigma\gamma_m I(x \in R_{jm})$$

Обновляем результат

$$F_m(x) = F_{m-1}(x) + \nu\Sigma\gamma_m I(x \in R_{jm})$$

На примере шага 1, где m = 1,
v = Learning rate

$$F_m(x) = F_{m-1}(x) + \nu\Sigma\gamma_m I(x \in R_{jm})$$

$$F_m(x) = 136.5$$

Обновляем результат

$$F_m(x) = F_{m-1}(x) + \nu \Sigma \gamma_m I(x \in R_{jm})$$

На примере шага 1, где m = 1, v = Learning rate

$$F_m(x) = F_{m-1}(x) + \nu \Sigma \gamma_m I(x \in R_{jm})$$

$$F_m(x) = 136.5 + \text{Learning Rate}$$

Обновляем результат

$$F_m(x) = F_{m-1}(x) + \nu \Sigma \gamma_m I(x \in R_{jm})$$

На примере шага 1, где m = 1, v = Learning rate

$$F_m(x) = F_{m-1}(x) + \nu \Sigma \gamma_m I(x \in R_{jm})$$

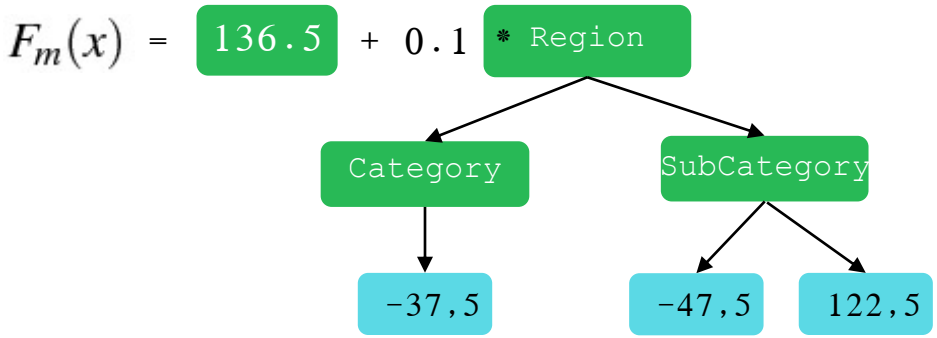
$$F_m(x) = 136.5 + 0.1$$

Обновляем результат

$$F_m(x) = F_{m-1}(x) + \nu\Sigma\gamma_m I(x \in R_{jm})$$

На примере шага 1, где m = 1, v = Learning rate

$$F_m(x) = F_{m-1}(x) + \nu\Sigma\gamma_m I(x \in R_{jm})$$



Обновляем результат

$$F_m(x) = F_{m-1}(x) + \nu \Sigma \gamma_m I(x \in R_{jm})$$

На примере шага 1, где m = 1, v = Learning rate

$$F_m(x) = F_{m-1}(x) + \nu \Sigma \gamma_m I(x \in R_{jm})$$

$$F_m(x) = 136.5 + 0.1 \; -37.5$$

Обновляем результат

$$F_m(x) = F_{m-1}(x) + \nu \Sigma \gamma_m I(x \in R_{jm})$$

На примере шага 1, где m = 1, v = Learning rate

$$F_m(x) = F_{m-1}(x) + \nu \Sigma \gamma_m I(x \in R_{jm})$$

$$F_m(x) = 136.5 + 0.1 \; -37.5 = 132.75$$

Обновляем результат

$$F_m(x) = F_{m-1}(x) + \nu\Sigma\gamma_m I(x \in R_{jm})$$

На примере шага 1, где m = 1, v = Learning rate

$$F_m(x) = F_{m-1}(x) + \nu\Sigma\gamma_m I(x \in R_{jm})$$

$$F_m(x) = 136.5 + 0.1 \; -37.5 = 132.75$$

⬇

Region	Category	Sub category	Param1	threshold	price	diff1	pred1
Владимирская область	Личные вещи	1	1	3	99	-37,5	132,75

Осталось сделать больше шагов

1. $Set(x_i, x_j)_{i=1}^{n}$ и функцию потерь $L(y_i, F(x))$
2. $F_0(x) = argmin\Sigma L(y_i, \gamma)$
3. Вычисление $r_{i,m} = -[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}]_{F(x)=F_{m-1}(x)}$
4. Обучение регрессии r_{jm} в цикле j = 1...m
5. В цикле j = 1...m вычисление $\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$
6. $F_m(x) = F_{m-1}(x) + \nu\Sigma\gamma_m I(x \in R_{jm})$

Посмотрим что мы посчитали

3

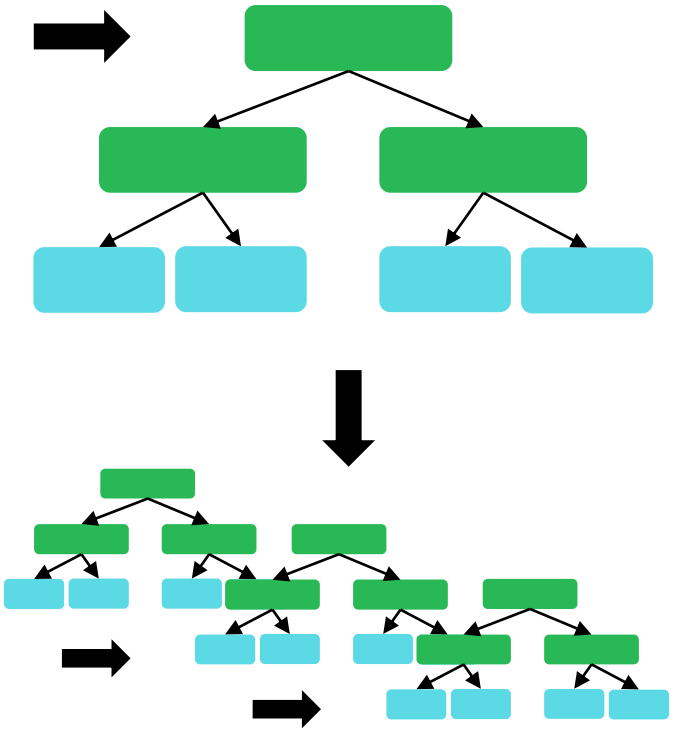
Градиент для классификации

В поиск класса

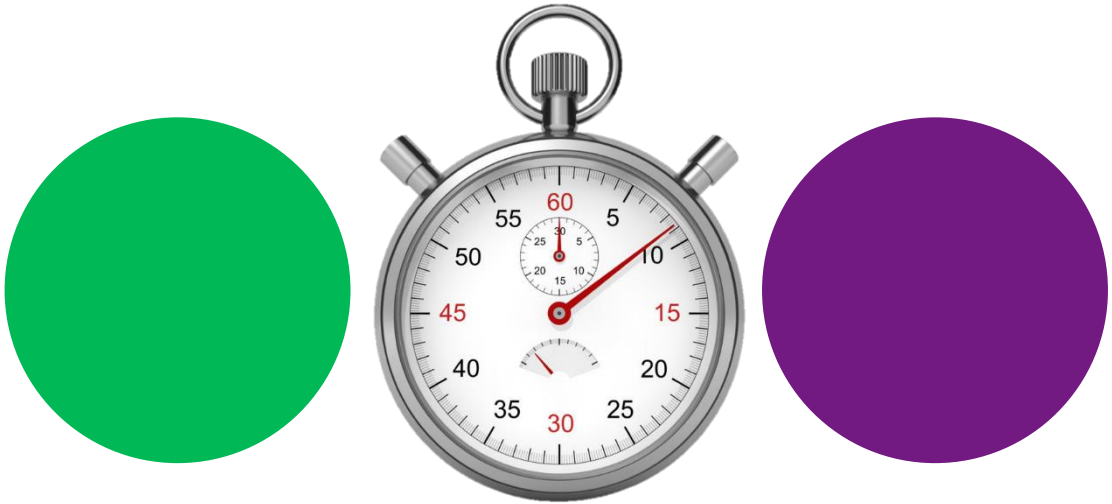
Region	Category	Sub category	Target
Владимирская область	Личные вещи	1	0
Волгоградская область	Личные вещи	2	0
Кировская область	Личные вещи	5	1

В поиск класса

Region	Category	Sub category	Target
Владимирская область	Личные вещи	1	0
Волгоградская область	Личные вещи	2	0
Кировская область	Личные вещи	5	1

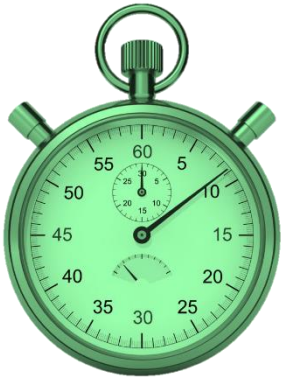


Нужен старт для расчета

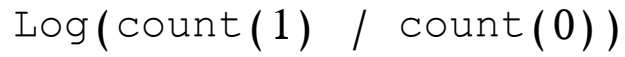


Нужен старт для расчета

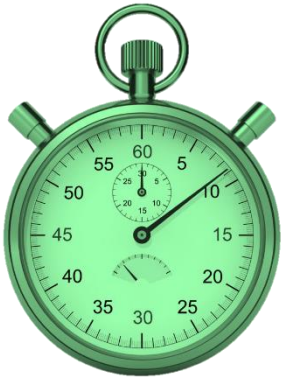
Log(odds of 1)



Нужен старт для расчета



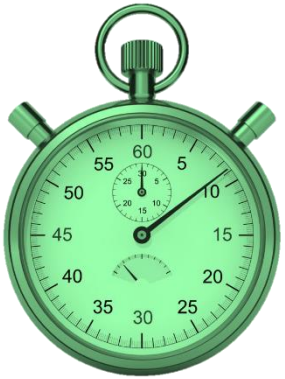
```
Log(count(1) / count(0))
```



Нужен старт для расчета

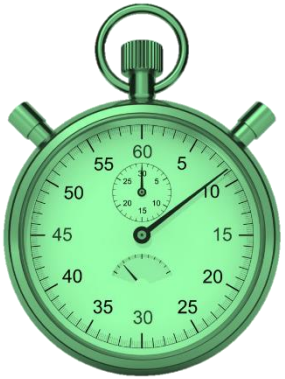
`Log(count(1) / count(0))`

$log(\frac{1}{2})$



Нужен старт для расчета

$$log(\frac{1}{2})$$



Вероятность 1

$$\frac{e^{log(\frac{|True|}{|False|})}}{1+e^{log(\frac{|True|}{|False|})}}$$

Вероятность 1

$$\frac{e^{log(\frac{|True|}{|False|})}}{1+e^{log(\frac{|True|}{|False|})}} = \frac{e^{log(\frac{1}{2})}}{1+e^{log(\frac{1}{2})}} = 0.33$$

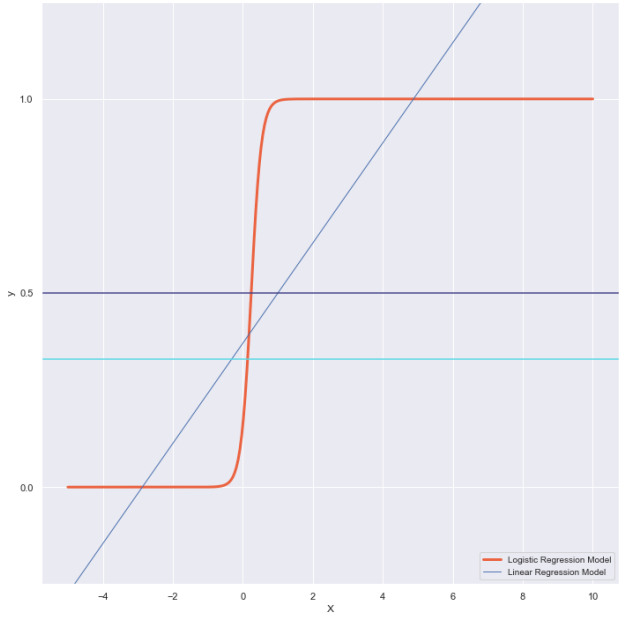
Вероятность 1

$$\frac{e^{log(\frac{|True|}{|False|})}}{1+e^{log(\frac{|True|}{|False|})}} = 0.3 = \text{порог для } 1$$

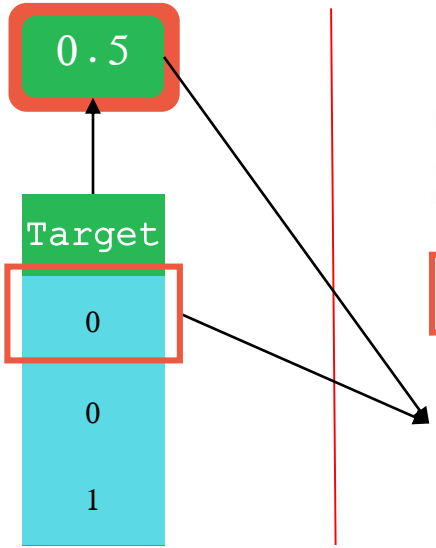
Вероятность 1

$$\frac{e^{log(\frac{|True|}{|False|})}}{1+e^{log(\frac{|True|}{|False|})}}$$

0.3 = порог для 1



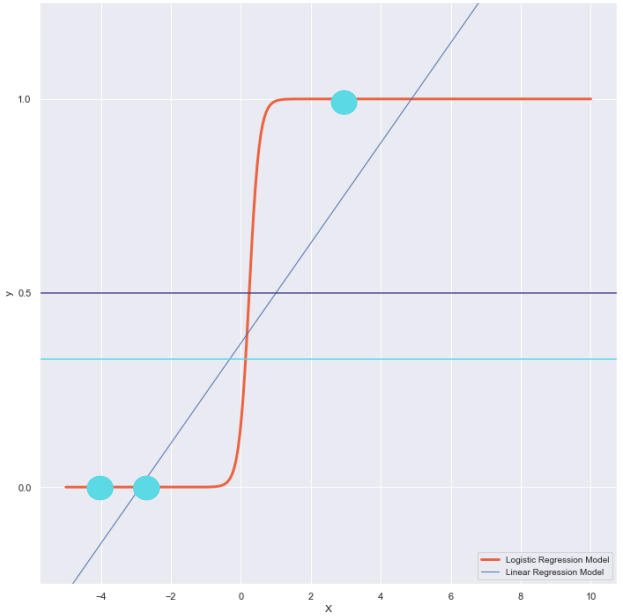
Вероятность 1



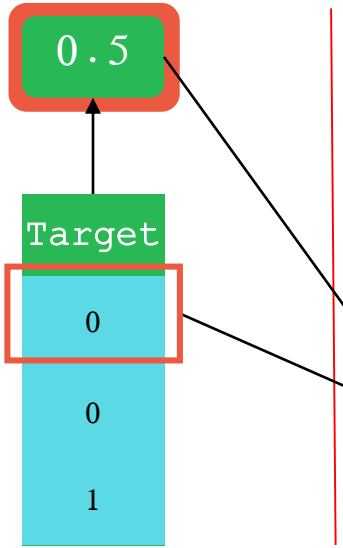
$$\frac{e^{\log\left(\frac{|True|}{|False|}\right)}}{1+e^{\log\left(\frac{|True|}{|False|}\right)}}$$

0.3 = порог для 1

0 − 0.5 = −0.5



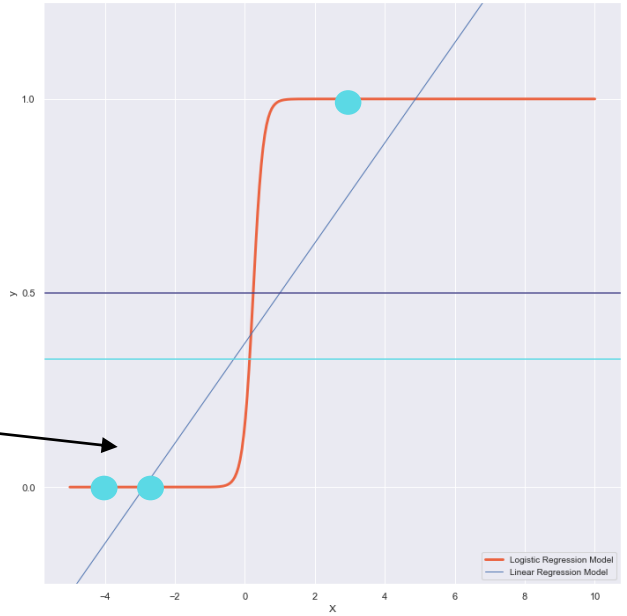
Вероятность 1



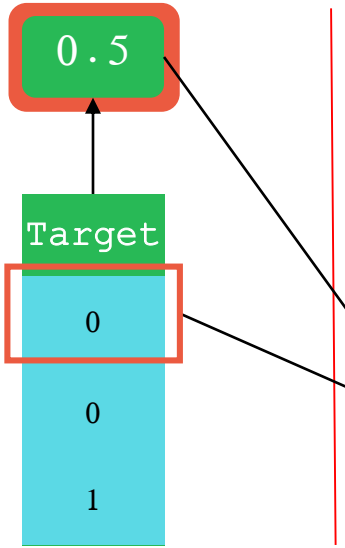
$$\frac{e^{log(\frac{|True|}{|False|})}}{1+e^{log(\frac{|True|}{|False|})}}$$

0.3 = порог для 1

0 − 0.5 = −0.5



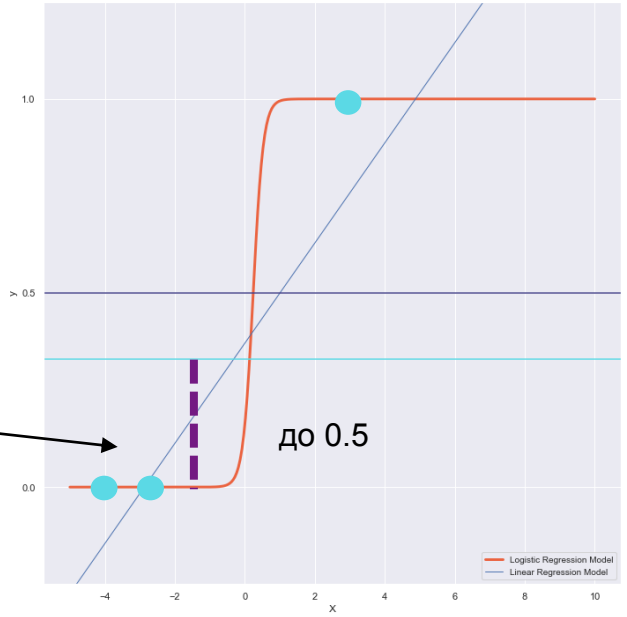
Вероятность 1



$$\frac{e^{\log\left(\frac{|True|}{|False|}\right)}}{1+e^{\log\left(\frac{|True|}{|False|}\right)}}$$

0.3 = порог для 1

0 − 0.5 = −0.5



Вероятность 1

Region	Category	Sub category	Target	diff
Владимирская область	Личные вещи	1	0	
Волгоградская область	Личные вещи	2	0	
Кировская область	Личные вещи	5	1	

(Observed – Predicted)

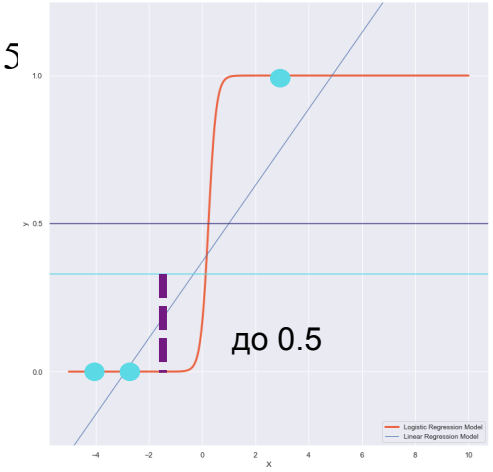
Вероятность 1

Region	Category	Sub category	Target	diff
Владимирская область	Личные вещи	1	0	-0,5
Волгоградская область	Личные вещи	2	0	-0,5
Кировская область	Личные вещи	5	1	0,5

$$(0 - 0.5)$$

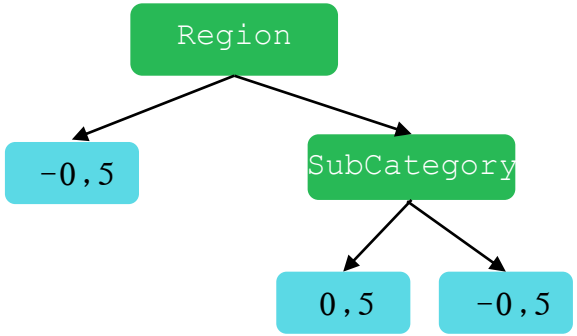
$$(0 - 0.5)$$

$$(1 - 0.5$$



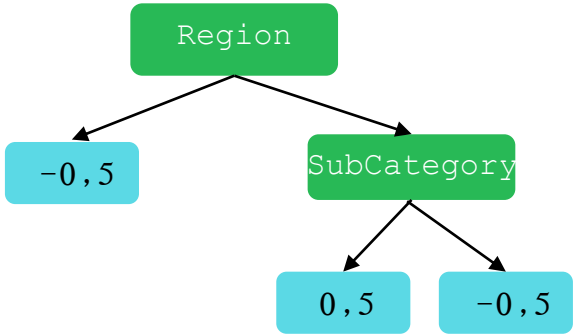
Вероятность 1

Region	Category	Sub category	Target	diff
Владимирская область	Личные вещи	1	0	−0,5
Волгоградская область	Личные вещи	2	0	−0,5
Кировская область	Личные вещи	5	1	0,5



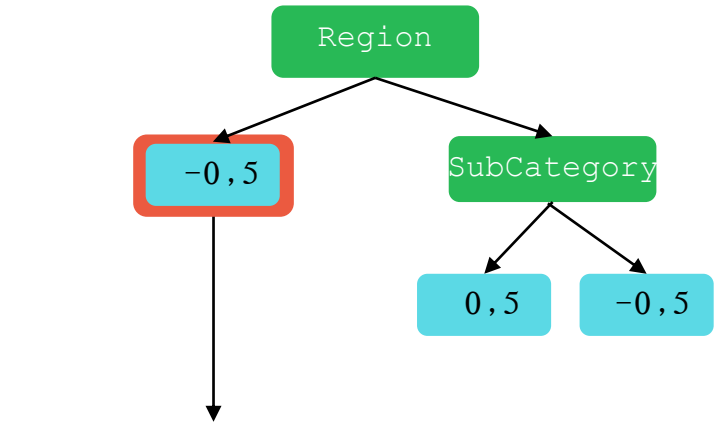
Вероятность 1

Region	Category	Sub category	Target	diff
Владимирская область	Личные вещи	1	0	−0,5
Волгоградская область	Личные вещи	2	0	−0,5
Кировская область	Личные вещи	5	1	0,5



$$\frac{\sum Diff_i}{\sum [PreviousProb_i * (1 - PreviousProb_i)]}$$

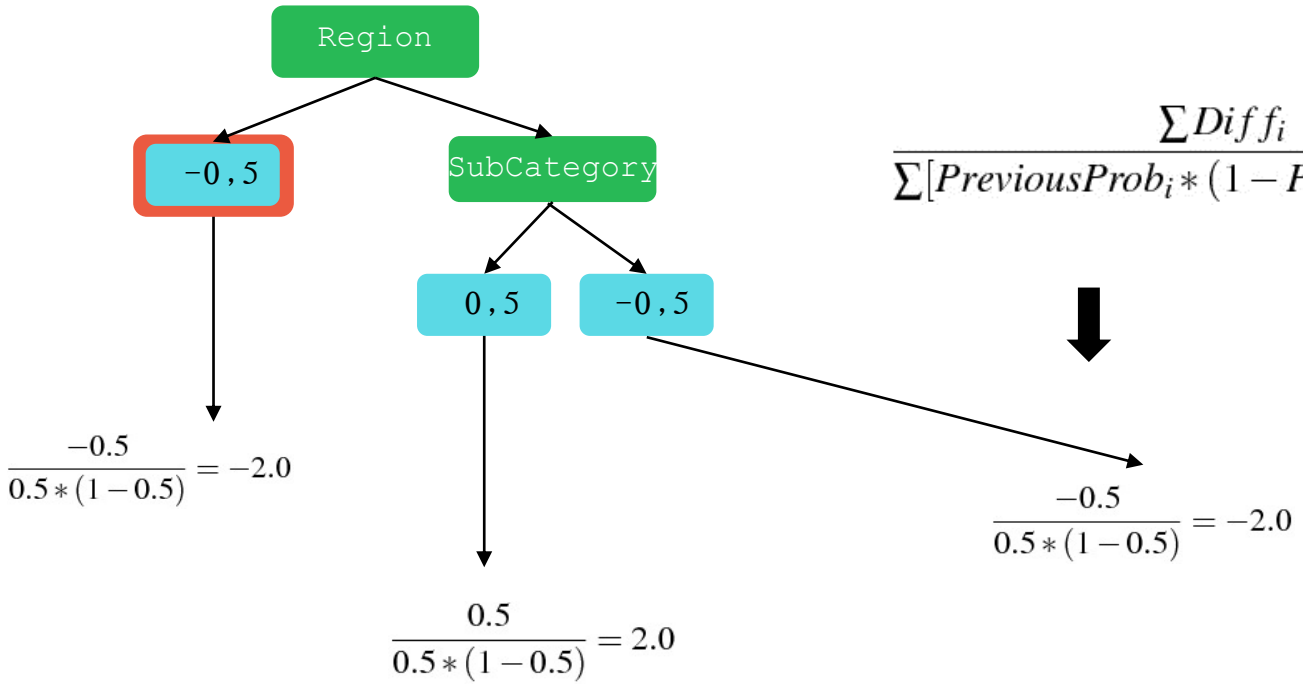
Вероятность



$$\frac{-0.5}{0.5*(1-0.5)} = -2.0$$

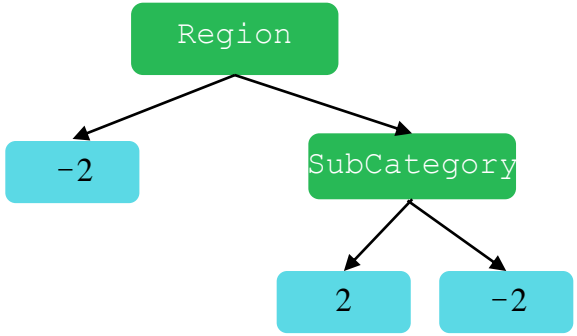
$$\frac{\sum Diff_i}{\sum[PreviousProb_i*(1-PreviousProb_i)]}$$

Вероятность



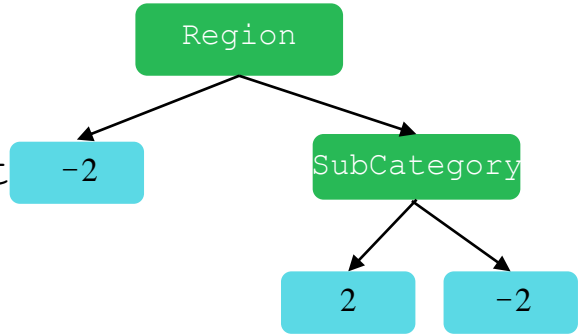
$$\frac{\sum Diff_i}{\sum[PreviousProb_i * (1 - PreviousProb_i)]}$$

Вероятность

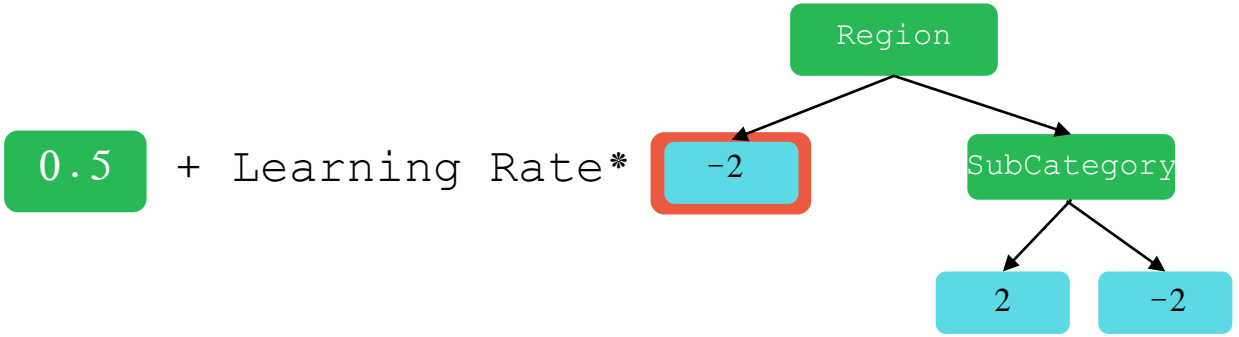


$$\frac{\sum Diff_i}{\sum [PreviousProb_i * (1 - PreviousProb_i)]}$$

0.5 + Learning Rat

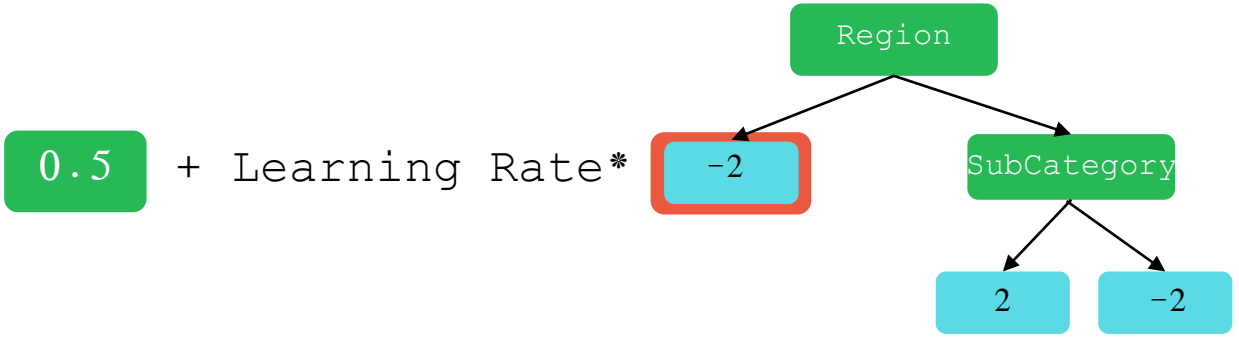


Вероятность



$$\text{Predict} = 0.5 + (0.1 * -2) = 0.3$$

Вероятность



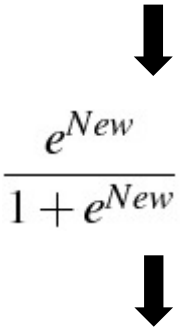
$$\text{Predict} = 0.5 + (0.1 * -2) = 0.3$$

$$\frac{e^{log(\frac{1}{2})}}{1+e^{log(\frac{1}{2})}} \rightarrow \frac{e^{0.3}}{1+e^{0.3}} = 0.57$$

Вероятность

Region	Category	Sub category	Target	Diff	Pred
Владимирская область	Личные вещи	1	0	-0,5	0,57
Волгоградская область	Личные вещи	2	0	-0,5	0,57
Кировская область	Личные вещи	5	1	0,5	0,66

Predict = 0.5 + (0.1 * Diff)



$$\frac{e^{New}}{1+e^{New}}$$

Predict Prob

Вероятность

Region	Category	Sub category	Target	Diff	Pred
Владимирская область	Личные вещи	1	0	-0,5	0,57
Волгоградская область	Личные вещи	2	0	-0,5	0,57
Кировская область	Личные вещи	5	1	0,5	0,66

$$Predict = 0.5 + (0.1 * Diff)$$

$$\frac{e^{New}}{1+e^{New}}$$

Predict Prob

Повторить много раз

Region	Category	Sub category	Target	pred
Владимирская область	Личные вещи	1	0	⬇
Волгоградская область	Личные вещи	2	0	⬇
Кировская область	Личные вещи	5	1	⬆

4

Больше деталей

Разобрали на примере

1. $Set(x_i, x_j)_{i=1}^{n}$ и функцию потерь $L(y_i, F(x))$
2. $F_0(x) = argmin\Sigma L(y_i, \gamma)$
3. Вычисление $r_{i,m} = -[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}]_{F(x)=F_{m-1}(x)}$
4. Обучение регрессии r_{jm} в цикле $j = 1 \ldots m$
5. В цикле $j = 1 \ldots m$ вычисление $\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$
6. $F_m(x) = F_{m-1}(x) + \nu\Sigma\gamma_m I(x \in R_{jm})$

Что изменилось

$$F_0(x) = \boxed{argmin \Sigma L(y_i, \gamma)}$$

⬇

для регрессии

$$\frac{1}{2}(Observed - Predicted)^2 = \frac{2}{2}(Observed - Predicted)x - 1 = -(Observed - Predicted)$$

Что изменилось

$$F_0(x) = argmin \Sigma L(y_i, \gamma)$$

⬇

для регрессии

$$\frac{1}{2}(Observed - Predicted)^2 = \frac{2}{2}(Observed - Predicted)x - 1 = -(Observed - Predicted)$$

⬇

для классификации

`Log(likelihood of Observed Data)`

Вероятность

$$F_0(x) = \boxed{argmin \Sigma L(y_i, \gamma)}$$

⬇

для регрессии

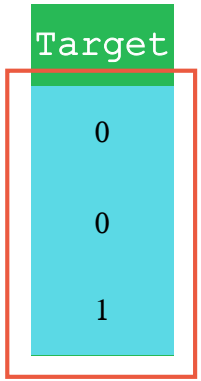
$$\frac{1}{2}(Observed - Predicted)^2 = \frac{2}{2}(Observed - Predicted)x - 1 = -(Observed - Predicted)$$

⬇

для классификации

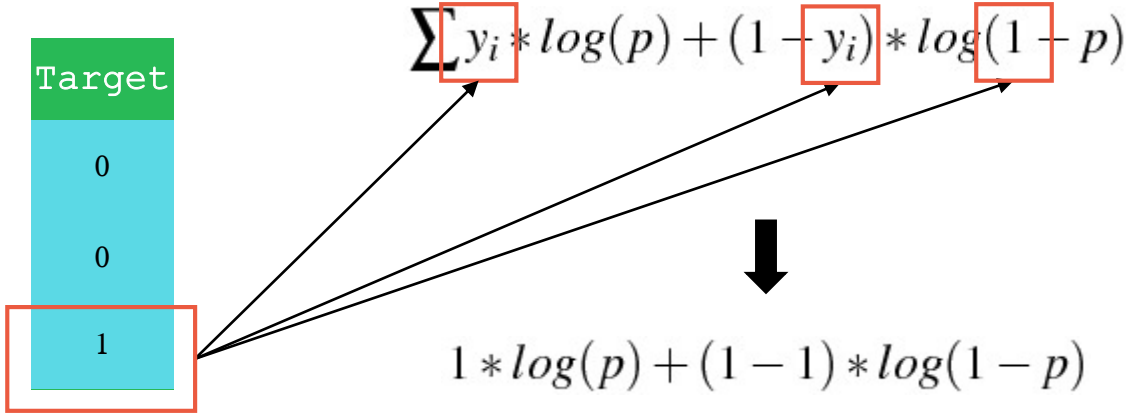
$$\sum y_i * log(p) + (1 - y_i) * log(1 - p) \qquad p = 0.5$$

Рассмотрим подробнее
log(likelihood)

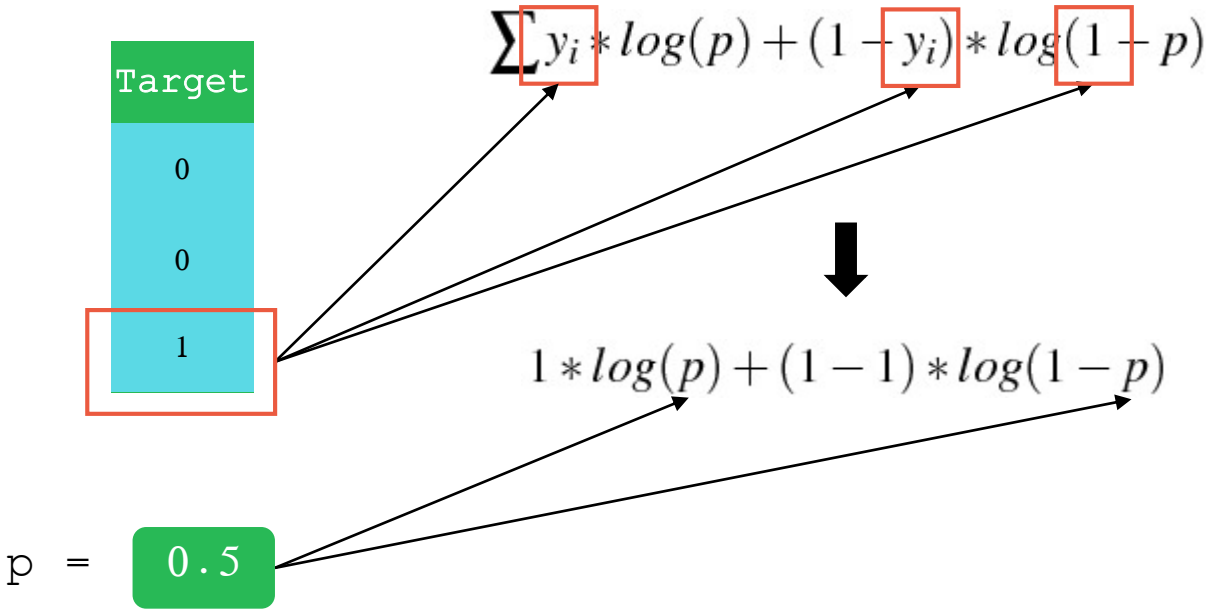


$$\sum y_i * log(p) + (1 - y_i) * log(1 - p)$$

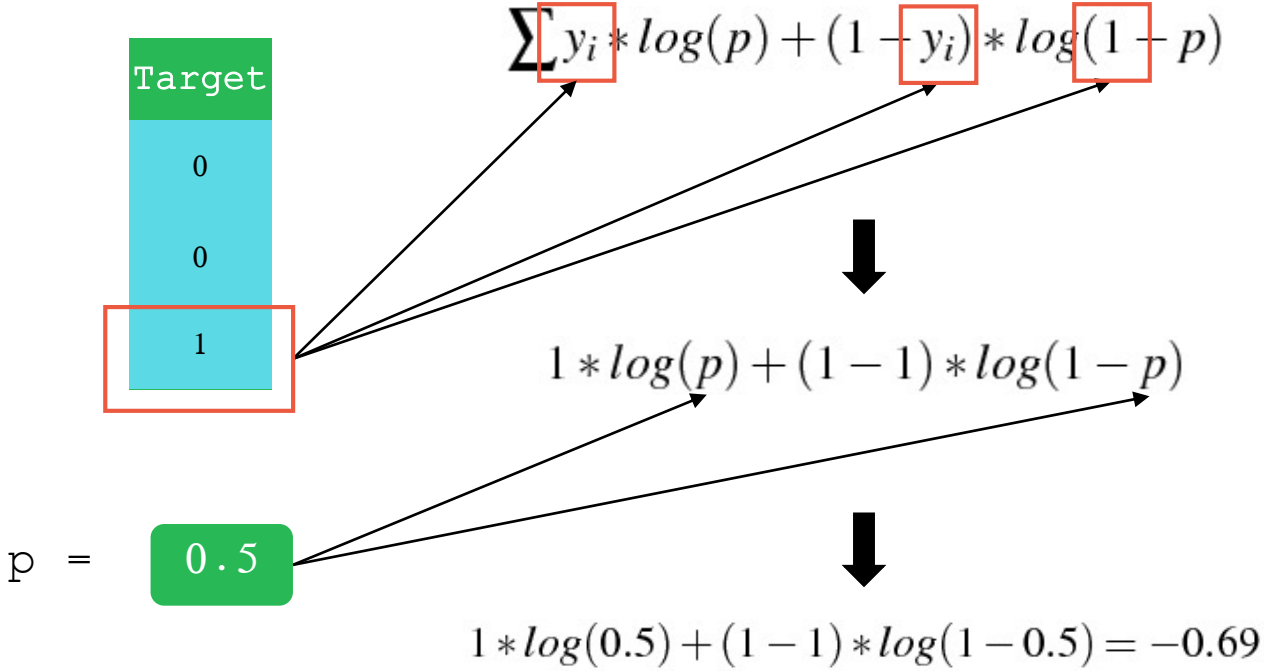
Рассмотрим подробнее
log(likelihood)



Рассмотрим подробнее
log(likelihood)

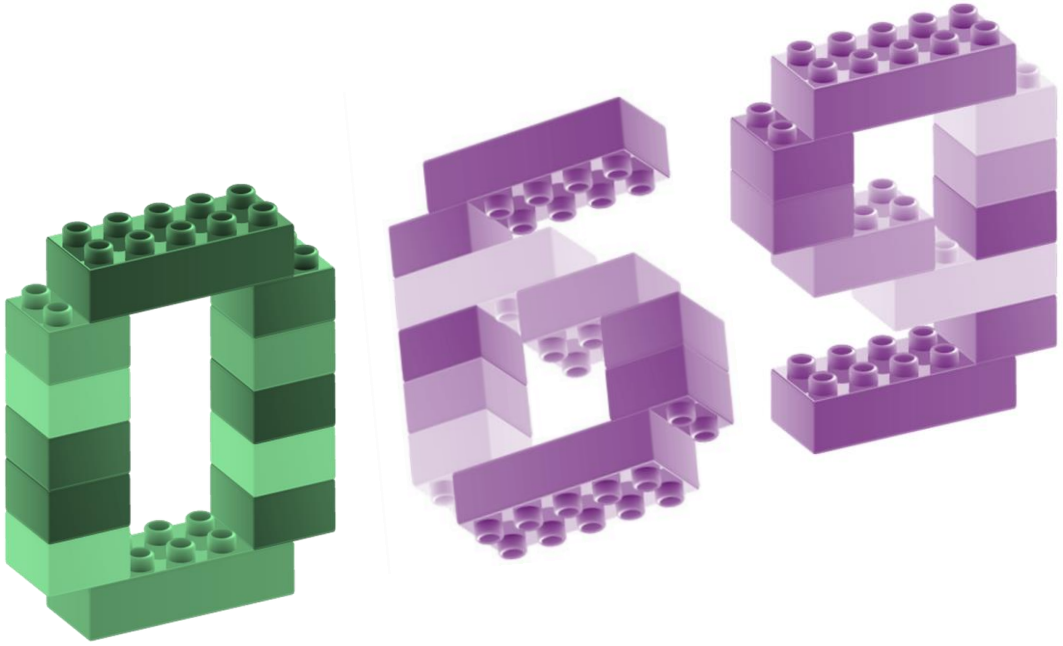


Рассмотрим подробнее
log(likelihood)



- 0.69?

$$\sum y_i * log(p) + (1 - y_i) * log(1 - p)$$



Упростим Loss

$$\sum y_i * log(p) + (1 - y_i) * log(1 - p) \quad \Rightarrow \quad -\sum y_i * log(p) + (1 - y_i) * log(1 - p)$$

Упростим Loss

$$\sum y_i * log(p) + (1 - y_i) * log(1 - p) \quad \Rightarrow \quad -\sum y_i * log(p) + (1 - y_i) * log(1 - p)$$

$$\Downarrow$$

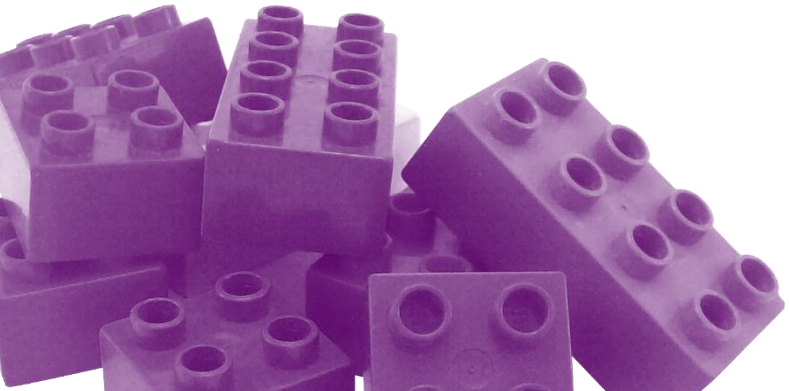
$$-[Observed * log(p) + (1 - Observed) * log(1 - p)]$$

Упростим Loss

$$\sum y_i * log(p) + (1 - y_i) * log(1 - p) \quad \Rightarrow \quad -\sum y_i * log(p) + (1 - y_i) * log(1 - p)$$

$$-[Observed * log(p) + (1 - Observed) * log(1 - p)]$$

$$-Observed * log(p) - (1 - Observed) * log(1 - p)$$



Упростим `Loss`

$$\sum y_i * log(p) + (1 - y_i) * log(1 - p) \quad \Rightarrow \quad -\sum y_i * log(p) + (1 - y_i) * log(1 - p)$$

$$\downarrow$$

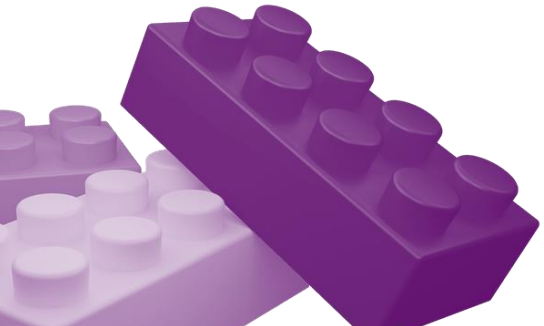
$$-[Observed * log(p) + (1 - Observed) * log(1 - p)]$$

$$\downarrow$$

$$-Observed * log(p) - (1 - Observed) * log(1 - p)$$

$$\downarrow$$

$$-Observed * log(p) - log(1 - p) + Observed * log(1 - p)$$



Упростим Loss

$$\sum y_i * log(p) + (1 - y_i) * log(1 - p) \quad \Rightarrow \quad -\sum y_i * log(p) + (1 - y_i) * log(1 - p)$$

$$\downarrow$$

$$-[Observed * log(p) + (1 - Observed) * log(1 - p)]$$

$$\downarrow$$

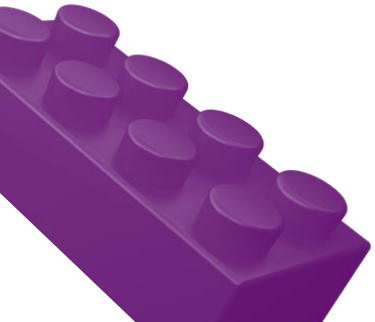
$$-Observed * log(p) - (1 - Observed) * log(1 - p)$$

$$\downarrow$$

$$-Observed * log(p) - log(1 - p) + Observed * log(1 - p)$$

$$\downarrow$$

$$-Observed * [log(p) - log(1 - p)] - log(1 - p)$$



Ещё несколько шагов

$$-Observed * [log(p) - log(1-p)] - log(1-p)$$

Ещё несколько шагов

$$-Observed * [log(p) - log(1-p)] - log(1-p)$$

$$log(p) - log(1-p) = \frac{log(p)}{log(1-p)}$$

Ещё несколько шагов

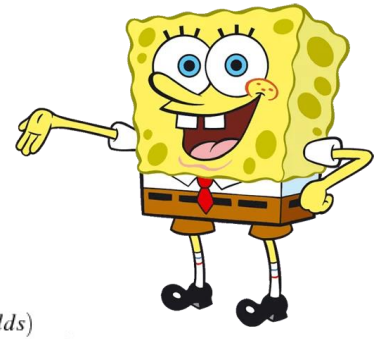
$$-Observed * [log(p) - log(1-p)] - log(1-p)$$

$$log(p) - log(1-p) = \frac{log(p)}{log(1-p)}$$

$$log(\frac{p}{1-p}) = \texttt{log(odds of 1)}$$

Ещё несколько шагов

$$-Observed * [log(p) - log(1-p)] - log(1-p)$$



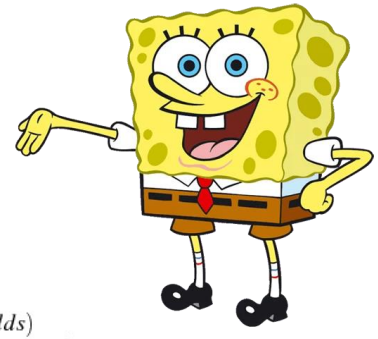
$$log(p) - log(1-p) = \frac{log(p)}{log(1-p)}$$

$$log(1-p) = log(1 - \frac{e^{log(odds)}}{1+e^{log(odds)}})$$

$$log(\frac{p}{1-p}) = \texttt{log(odds of 1)}$$

Ещё несколько шагов

$$-Observed * [log(p) - log(1-p)] - log(1-p)$$



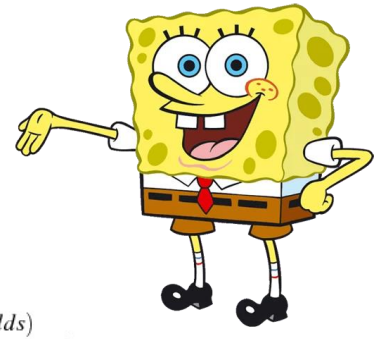
$$log(p) - log(1-p) = \frac{log(p)}{log(1-p)}$$

$$log(\frac{p}{1-p}) = \texttt{log(odds of 1)}$$

$$log(1-p) = log(1 - \frac{e^{log(odds)}}{1+e^{log(odds)}})$$

Ещё несколько шагов

$$-Observed * [log(p) - log(1-p)] - log(1-p)$$



$$log(p) - log(1-p) = \frac{log(p)}{log(1-p)}$$

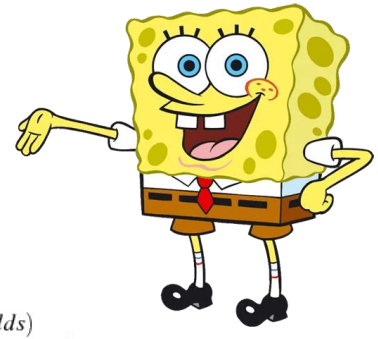
$$log(\frac{p}{1-p}) = \texttt{log(odds of 1)}$$

$$log(1-p) = log(1 - \frac{e^{log(odds)}}{1+e^{log(odds)}})$$

$$log(\frac{1+e^{log(odds)}}{1+e^{log(odds)}} - \frac{e^{log(odds)}}{1+e^{log(odds)}})$$

Ещё несколько шагов

$$-Observed * [log(p) - log(1-p)] - log(1-p)$$



$$log(p) - log(1-p) = \frac{log(p)}{log(1-p)}$$

$$log(\frac{p}{1-p}) = \texttt{log(odds of 1)}$$

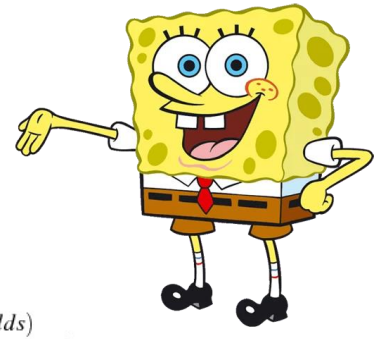
$$log(1-p) = log(1 - \frac{e^{log(odds)}}{1+e^{log(odds)}})$$

$$log(\frac{1+e^{log(odds)}}{1+e^{log(odds)}} - \frac{e^{log(odds)}}{1+e^{log(odds)}})$$

$$log(\frac{1}{1+e^{log(odds)}}) = log(1) - log(1+e^{log(odds)})$$

Ещё несколько шагов

$$-Observed * [log(p) - log(1-p)] - log(1-p)$$



$$log(p) - log(1-p) = \frac{log(p)}{log(1-p)}$$

$$log(\frac{p}{1-p}) = \texttt{log(odds of 1)}$$

$$log(1-p) = log(1 - \frac{e^{log(odds)}}{1+e^{log(odds)}})$$

$$log(\frac{1+e^{log(odds)}}{1+e^{log(odds)}} - \frac{e^{log(odds)}}{1+e^{log(odds)}})$$

$$log(\frac{1}{1+e^{log(odds)}}) = log(1) - log(1+e^{log(odds)})$$

$$-log(1+e^{log(odds)})$$

Ещё несколько шагов

$$-Observed * [log(p) - log(1-p)] - log(1-p)$$

$$log(\frac{p}{1-p})$$ = log(odds of 1)

$$-log(1+e^{log(odds)})$$

Ещё несколько шагов

$$-Observed * [log(p) - log(1-p)] - log(1-p)$$

$$log(\frac{p}{1-p}) = \texttt{log(odds of 1)}$$

$$-log(1+e^{log(odds)})$$

$$-Observed * log(odds) + log(1+e^{log(odds)})$$

Возьмем производную (от Loss

$$\frac{\partial}{\partial log(odds)} - Observed * log(odds) + log(1 + e^{log(odds)})$$

Возьмем производную от Loss

$$\frac{\partial}{\partial log(odds)} - Observed * log(odds) + log(1 + e^{log(odds)})$$

- Observed

Возьмем производную от Loss

$$\frac{\partial}{\partial log(odds)} - Observed * log(odds) + log(1 + e^{log(odds)})$$

⬇ ⬇

- Observed

$$\frac{1}{1 + e^{log(odds)}}$$

Возьмем производную от Loss

$$\frac{\partial}{\partial log(odds)} - Observed * log(odds) + log(1 + e^{log(odds)})$$

$$-Observed + \frac{1}{1 + e^{log(odds)}} * e^{log(odds)}$$

Возьмем производную от Loss

$$\frac{\partial}{\partial log(odds)} - Observed * log(odds) + log(1 + e^{log(odds)})$$

$$-Observed + \frac{1}{1 + e^{log(odds)}} * e^{log(odds)}$$

Возьмем производную от Loss

$$\frac{\partial}{\partial log(odds)} - Observed * log(odds) + log(1 + e^{log(odds)})$$

$$-Observed + \frac{1}{1 + e^{log(odds)}} * e^{log(odds)}$$

$$-Observed + \frac{e^{log(odds)}}{1 + e^{log(odds)}}$$

5

Градиентный спуск для классификации

Argmin <- минимизируем потерю

$$F_0(x) = argmin \Sigma L(y_i, \gamma)$$

price
99
99
89
259

$$-0 * log(odds) + log(1 + e^{log(odds)})$$

$$-0 * log(odds) + log(1 + e^{log(odds)})$$

$$-0 * log(odds) + log(1 + e^{log(odds)})$$

Argmin <- минимизируем потерю

$$F_0(x) = argmin \Sigma L(y_i, \gamma)$$

price
99
99
89
259

$$-0 * log(odds) + log(1 + e^{log(odds)}) \rightarrow -0 + \frac{e^{log(odds)}}{1 + e^{log(odds)}}$$

$$-0 * log(odds) + log(1 + e^{log(odds)}) \rightarrow -0 + \frac{e^{log(odds)}}{1 + e^{log(odds)}}$$

$$-0 * log(odds) + log(1 + e^{log(odds)}) \rightarrow -0 + \frac{e^{log(odds)}}{1 + e^{log(odds)}}$$

Argmin <- минимизируем потерю

$$F_0(x) = argmin \Sigma L(y_i, \gamma)$$

price
99
99
89
259

$$-0 * log(odds) + log(1 + e^{log(odds)}) \rightarrow -0 + \frac{e^{log(odds)}}{1 + e^{log(odds)}} \rightarrow -0 + p$$

$$-0 * log(odds) + log(1 + e^{log(odds)}) \rightarrow -0 + \frac{e^{log(odds)}}{1 + e^{log(odds)}} \rightarrow -0 + p$$

$$-0 * log(odds) + log(1 + e^{log(odds)}) \rightarrow -0 + \frac{e^{log(odds)}}{1 + e^{log(odds)}} \rightarrow -1 + p$$

Argmin <- минимизируем потерю

$$F_0(x) = argmin \Sigma L(y_i, \gamma)$$

price
99
99
89
259

$$-0 + p + (-0) + p + (-1) + p = 0$$

Argmin <- минимизируем потерю

$$F_0(x) = argmin \Sigma L(y_i, \gamma)$$

price
99
99
89
259

⬇

$$-0 + p + (-0) + p + (-1) + p = 0$$

⬇

$$-1 + 3p$$

$$3p = 1$$

$$p = 1/3 = 0.33$$

Argmin <- минимизируем потерю

$$F_0(x) = argmin \Sigma L(y_i, \gamma)$$

price
99
99
89
259

⬇

$$-0 + p + (-0) + p + (-1) + p = 0$$

⬇

$$p = \frac{1}{3} \quad => \quad log(\frac{p}{1-p})$$

Argmin <- минимизируем потерю

$$F_0(x) = argmin \Sigma L(y_i, \gamma)$$

price
99
99
89
259

⬇

$$-0 + p + (-0) + p + (-1) + p = 0$$

⬇

$$p = \frac{1}{3} \Rightarrow log(\frac{p}{1-p})$$

⬇

$$log(\frac{\frac{1}{3}}{1-\frac{1}{3}}) \Rightarrow -0.69$$

Argmin <- минимизируем потерю

$$F_0(x) = argmin \Sigma L(y_i, \gamma)$$

price
99
99
89
259

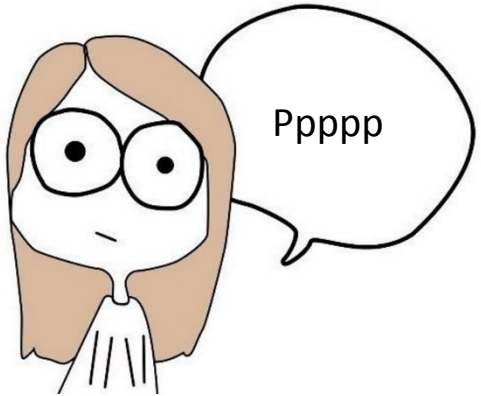
$$F_0(x) = -0.69$$

Теперь разберём

1. $Set(x_i, x_j)_{i=1}^{n}$ и функцию потерь $L(y_i, F(x))$
2. $F_0(x) = argmin\Sigma L(y_i, \gamma)$
3. Вычисление $r_{i,m} = -[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}]_{F(x)=F_{m-1}(x)}$
4. Обучение регрессии r_{jm} в цикле j = 1...m
5. В цикле j = 1...m вычисление $\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$
6. $F_m(x) = F_{m-1}(x) + \nu\Sigma\gamma_m I(x \in R_{jm})$

Вычислим R

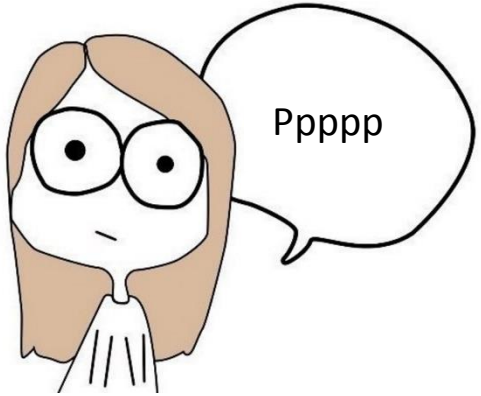
$$r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$



Вычислим R

$$r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

$$-Observed + \frac{e^{log(odds)}}{1 + e^{log(odds)}}$$

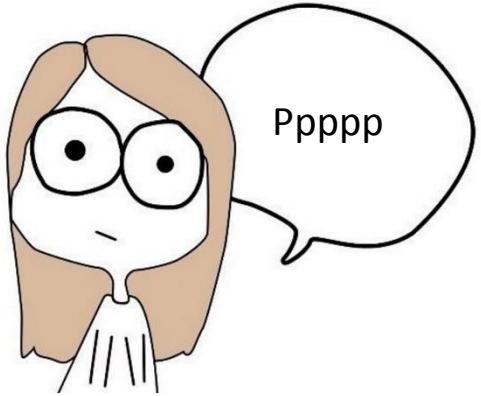


Вычислим R

$$r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

$$-Observed + \frac{e^{log(odds)}}{1 + e^{log(odds)}}$$

-0.69



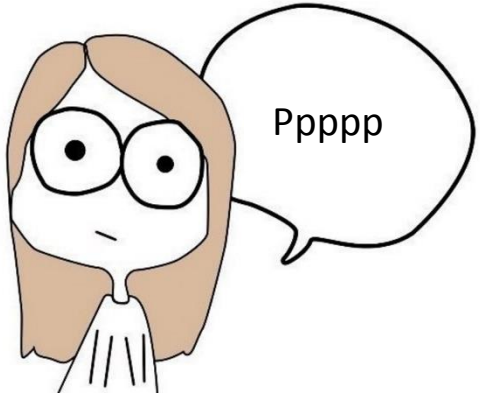
Вычислим R

$$r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

$$-Observed + \frac{e^{log(odds)}}{1 + e^{log(odds)}}$$

$$Observed - \frac{e^{-0.69}}{1 + e^{-0.69}}$$

- 0.69



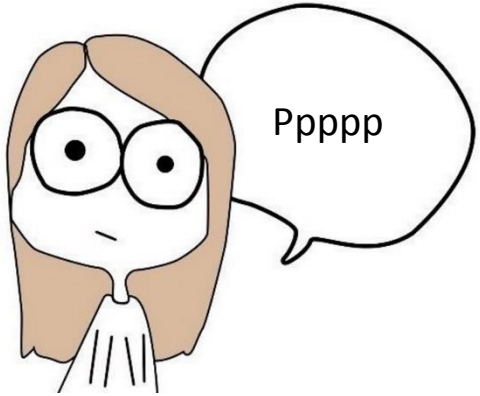
Вычислим R

$$r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

$$-Observed + \frac{e^{log(odds)}}{1 + e^{log(odds)}}$$

$$Observed - \frac{e^{-0.69}}{1 + e^{-0.69}}$$

- 0.69



Observed - p = Observed - 0.33

Вычислим R

$$r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

Region	Category	Sub category	Target	R	
Владимирская область	Личные вещи	1	0	-0,33	← Observed - 0.33
Волгоградская область	Личные вещи	2	0	-0,33	← Observed - 0.33
Кировская область	Личные вещи	5	1	0,67	← Observed - 0.33

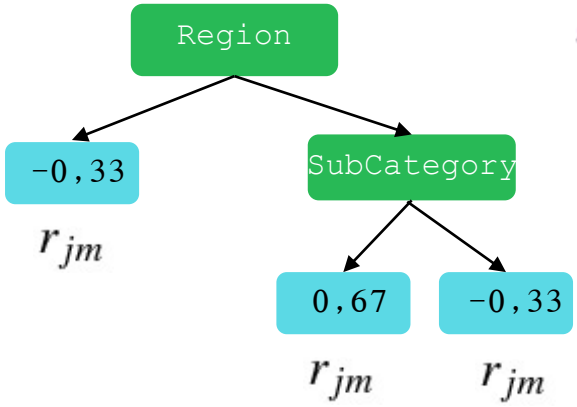
Вычислим R

$$r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

Region	Category	Sub category	Target	R	
Владимирская область	Личные вещи	1	0	-0,33	← 0 - 0.33
Волгоградская область	Личные вещи	2	0	-0,33	← 0 - 0.33
Кировская область	Личные вещи	5	1	0,67	← 1 - 0.33

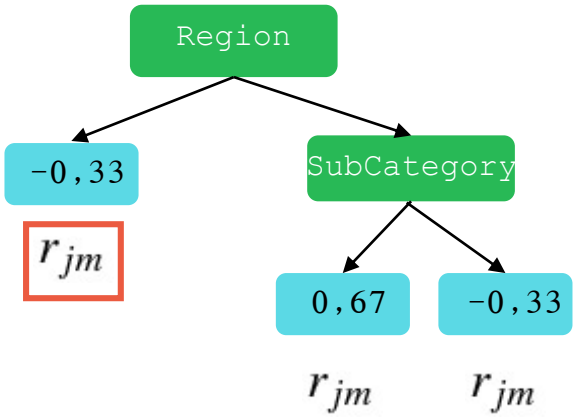
Обучение регрессии на дереве

Region	Category	Sub category	Target	R
Владимирская область	Личные вещи	1	0	−0,33
Волгоградская область	Личные вещи	2	0	−0,33
Кировская область	Личные вещи	5	1	0,67



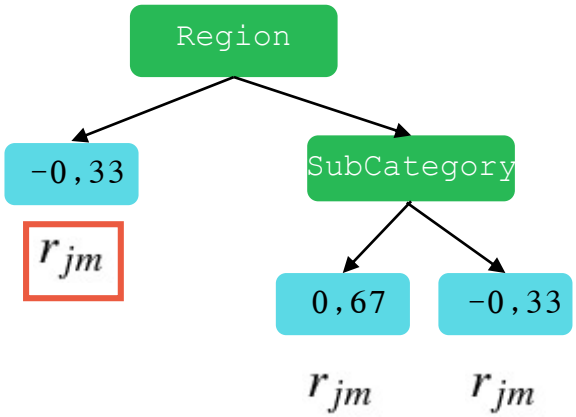
Вычисление

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



Вычисление

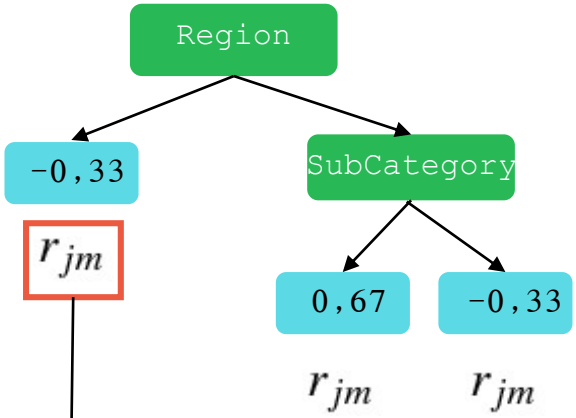
$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

Вычисление

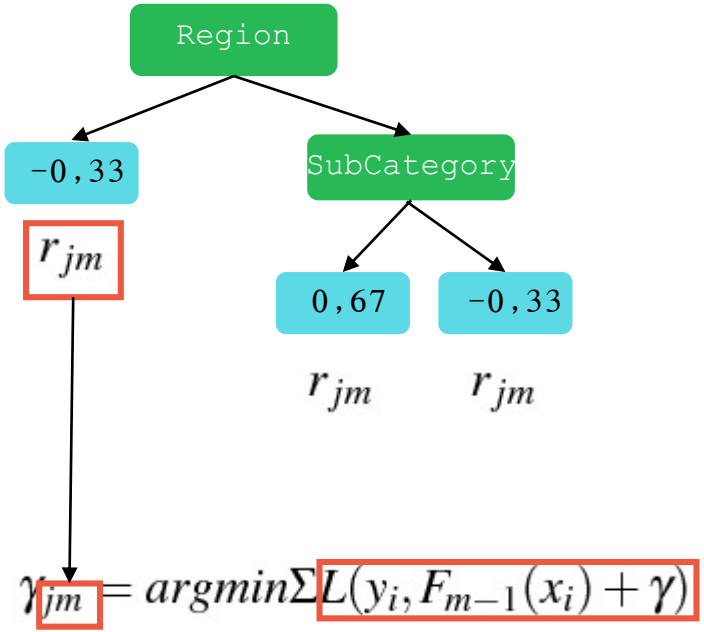
$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

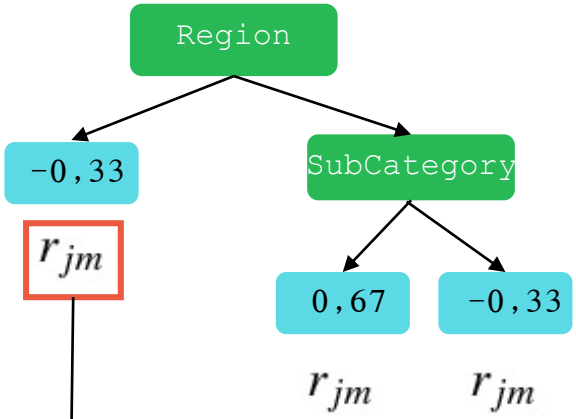
Вычисление

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



Вычисление

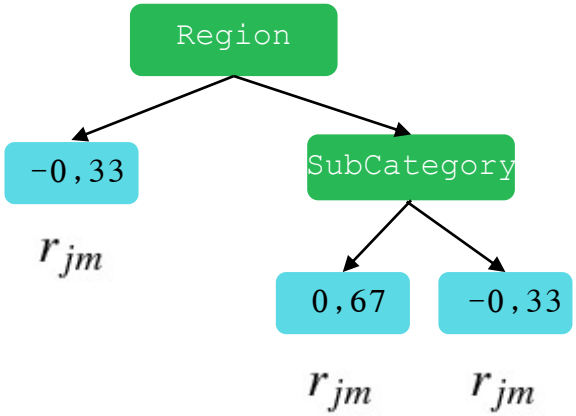
$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma) \rightarrow -Observed * log(odds) + log(1 + e^{log(odds)})$$

Вычисление

$$\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



r_{jm} r_{jm} r_{jm}

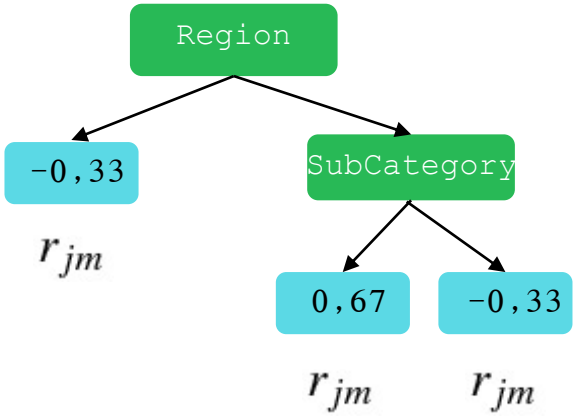
$$\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$-Observed * log(odds) + log(1 + e^{log(odds)})$$

$$argmin\Sigma - y_i * [F_{m-1}(x_i) + \gamma] + log(1 + e^{F_{m-1}(x_i)+\gamma})$$

Вычисление

$$\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



r_{jm} r_{jm} r_{jm}

$$\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

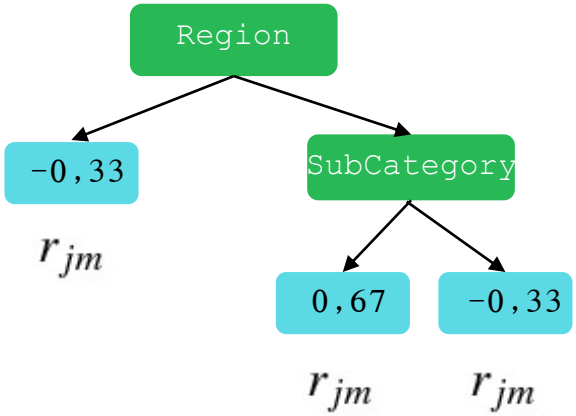
$$-Observed * log(odds) + log(1 + e^{log(odds)})$$

$$argmin\Sigma - y_i * [F_{m-1}(x_i) + \gamma] + log(1 + e^{F_{m-1}(x_i)+\gamma})$$

-0,33

Вычисление

$$\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



r_{jm} r_{jm} r_{jm}

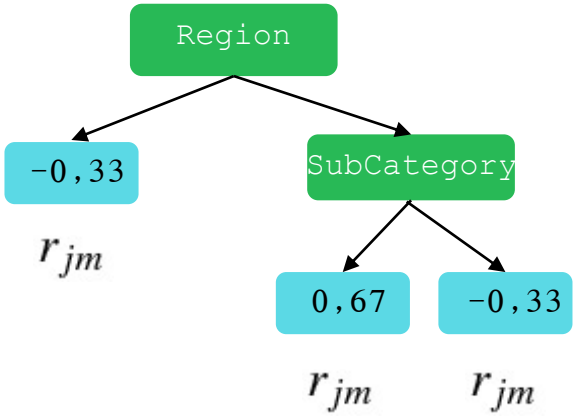
$$\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$-Observed * log(odds) + log(1 + e^{log(odds)})$$

$$argmin\Sigma - y_i * [F_{m-1}(x_i) + \gamma] + log(1 + e^{F_{m-1}(x_i)+\gamma})$$

Вычисление

$$\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



r_{jm} r_{jm} r_{jm}

$$\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

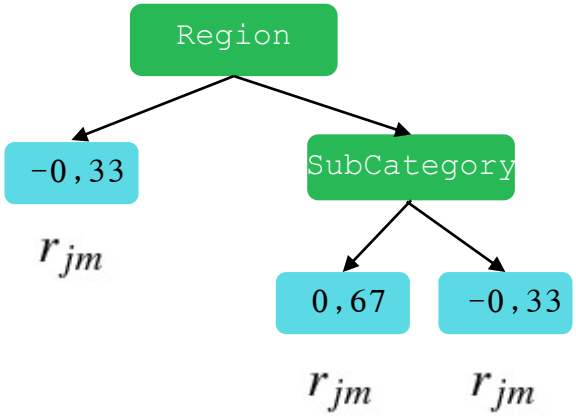
$$-Observed * log(odds) + log(1 + e^{log(odds)})$$

$$argmin\Sigma - y_i * [F_{m-1}(x_i) + \gamma] + log(1 + e^{F_{m-1}(x_i)+\gamma})$$

$$\frac{\partial}{\partial \gamma} L(y_i, F_{m-1}(x_i) + \gamma) \approx \frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i)) + \frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i))\gamma$$

Вычисление

$$\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$-Observed * log(odds) + log(1 + e^{log(odds)})$$

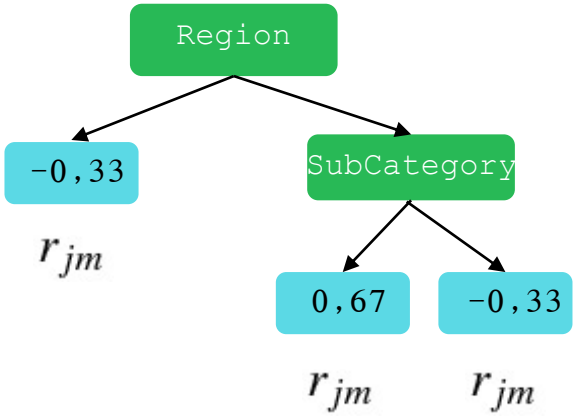
$$argmin\Sigma - y_i * [F_{m-1}(x_i) + \gamma] + log(1 + e^{F_{m-1}(x_i)+\gamma})$$

$$\frac{\partial}{\partial \gamma} L(y_i, F_{m-1}(x_i) + \gamma) \approx \frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i)) + \frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i))\gamma$$

$$\frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i))\gamma = \frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i))$$

Вычисление

$$\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$-Observed * log(odds) + log(1 + e^{log(odds)})$$

$$argmin\Sigma - y_i * [F_{m-1}(x_i) + \gamma] + log(1 + e^{F_{m-1}(x_i)+\gamma})$$

$$\downarrow$$

$$\frac{\partial}{\partial \gamma} L(y_i, F_{m-1}(x_i) + \gamma) \approx \frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i)) + \frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i))\gamma$$

$$\downarrow$$

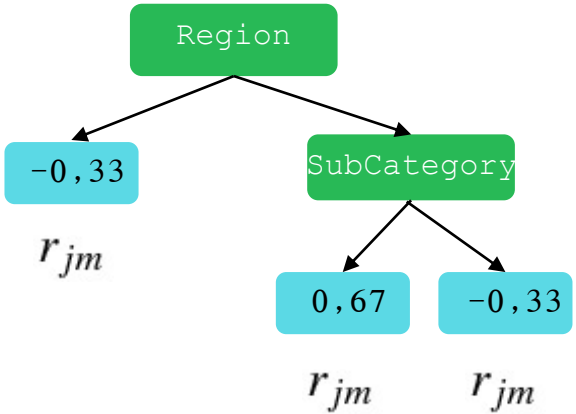
$$\frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i))\gamma = \frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i))$$

$$\downarrow$$

$$\gamma = \frac{-\frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i))}{\frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i))}$$

Вычисление

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



r_{jm}

r_{jm} r_{jm}

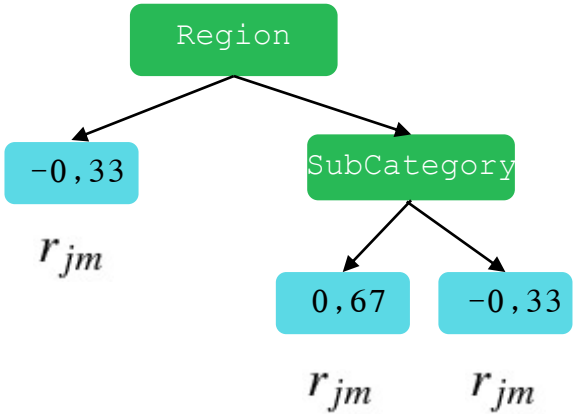
$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$-Observed * log(odds) + log(1 + e^{log(odds)})$$

$$\gamma = \frac{-\frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i))}{\frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i))}$$

Вычисление

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



r_{jm} r_{jm} r_{jm}

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

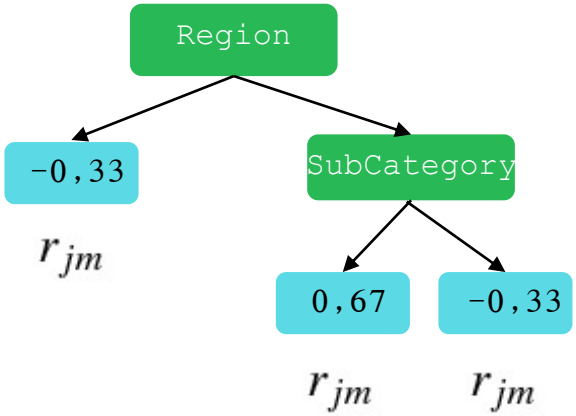
$$-Observed * log(odds) + log(1 + e^{log(odds)})$$

Observed - p

$$\gamma = \frac{-\frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i))}{\frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i))}$$

Вычисление

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



r_{jm} r_{jm} r_{jm}

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

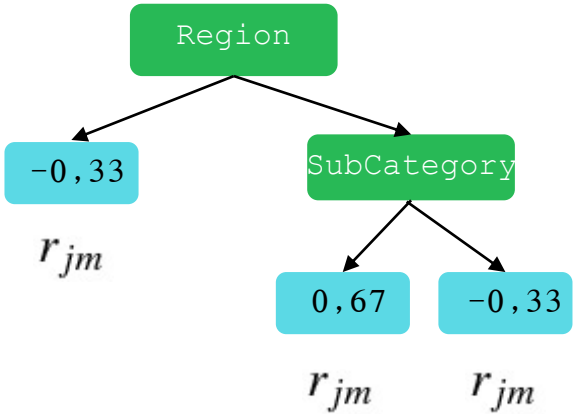
$$-Observed * log(odds) + log(1 + e^{log(odds)})$$

$$\gamma = \frac{-\frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i))}{\frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i))}$$

$$\frac{\partial^2}{\partial log(odds)^2} - Observed * log(odds) + log(1 + e^{log(odds)})$$

Вычисление

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$-Observed * log(odds) + log(1 + e^{log(odds)})$$

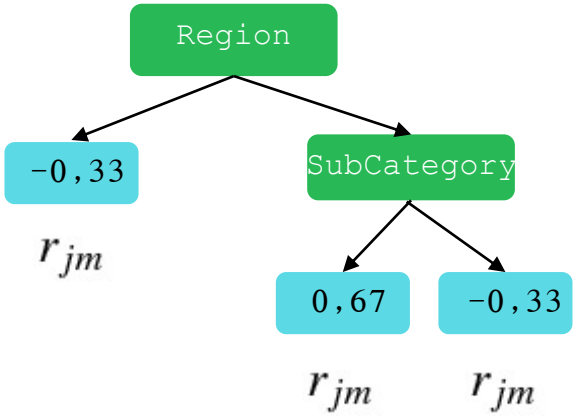
$$\gamma = \frac{-\frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i))}{\frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i))}$$

↓

p * (1 - p)

Вычисление

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



r_{jm} r_{jm} r_{jm}

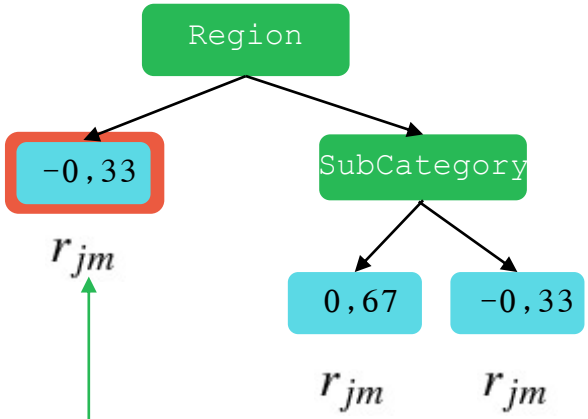
$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$-Observed * log(odds) + log(1 + e^{log(odds)})$$

$$\frac{Observed - p}{p * (1 - p)}$$

Вычисление

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



r_{jm} r_{jm} r_{jm}

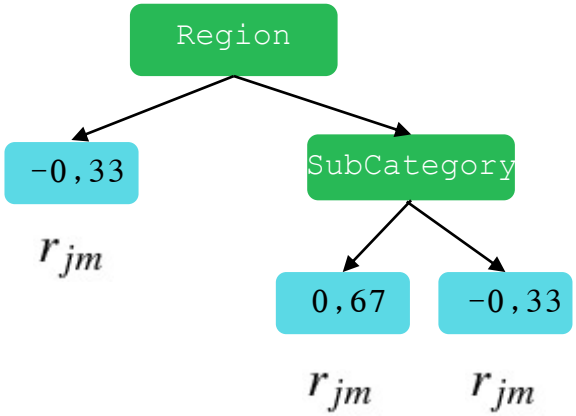
$$\gamma_{1,1} = \frac{-0.33}{-0.69 * (1 - (-0.69))}$$

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$-Observed * log(odds) + log(1 + e^{log(odds)})$$

Вычисление

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



r_{jm} r_{jm} r_{jm}

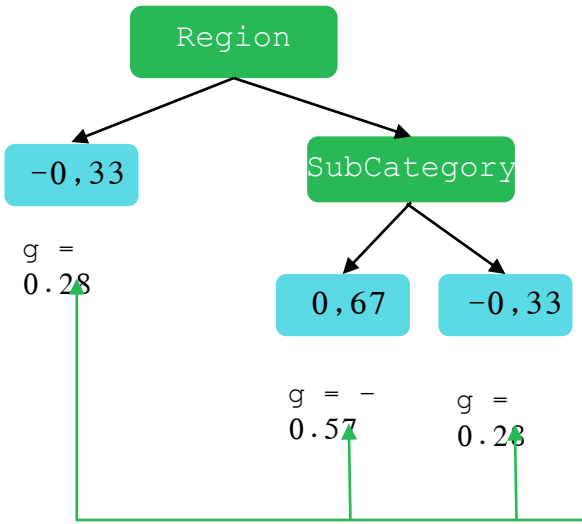
$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$-Observed * log(odds) + log(1 + e^{log(odds)})$$

$$\gamma_{1,1} = \frac{-0.33}{-0.69 * (1 - (-0.69))} = 0.28$$

Вычисление

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\frac{Observed - p}{p * (1 - p)}$$

$$\gamma_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$-Observed * log(odds) + log(1 + e^{log(odds)})$$

Gamma для всех

$$\frac{Observed - p}{p * (1 - p)}$$

Region	Category	Sub category	Target	R	G
Владимирская область	Личные вещи	1	0	-0,33	
Волгоградская область	Личные вещи	2	0	-0,33	
Кировская область	Личные вещи	5	1	0,67	

← 0.28

← 0.28

← -0.57

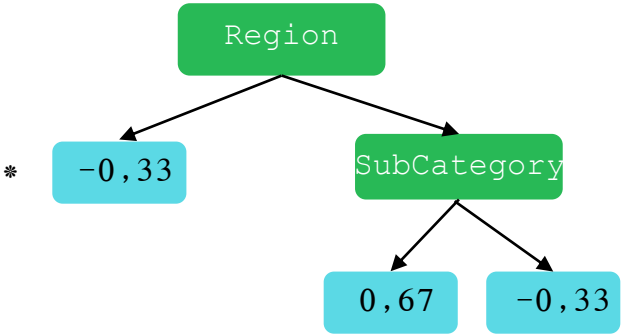
Обновляем F

$$F_m(x) = F_{m-1}(x) + \nu \Sigma \gamma_m I(x \in R_{jm})$$

Обновляем F

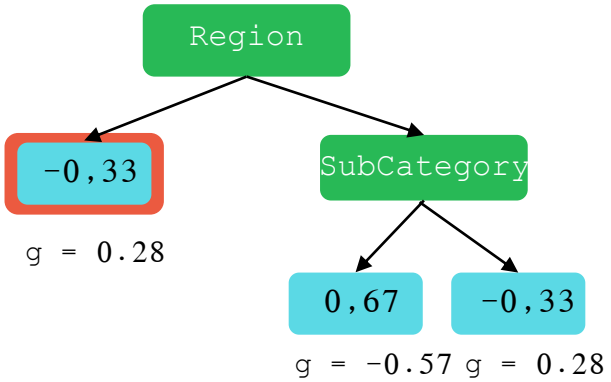
$$F_m(x) = F_{m-1}(x) + \nu \Sigma \gamma_m I(x \in R_{jm})$$

0.5 + Learning Rate *



Обновляем F

$$F_m(x) = F_{m-1}(x) + \nu \Sigma \gamma_m I(x \in R_{jm})$$



F = 0.5 + (0.1 * 0.28) = 0.58

Новый предикт для всех

Region	Category	Sub category	Target	Predict
Владимирская область	Личные вещи	1	0	0,58
Волгоградская область	Личные вещи	2	0	0,58
Кировская область	Личные вещи	5	1	0,44

$\leftarrow 0.5 + (0.1 * 0.28) = 0.58$

$\leftarrow 0.5 + (0.1 * 0.28) = 0.58$

$\leftarrow 0.5 + (0.1 * -0.57) = 0.44$

Завершили цикл

1. $Set(x_i, x_j)_{i=1}^{n}$ и функцию потерь $L(y_i, F(x))$
2. $F_0(x) = argmin\Sigma L(y_i, \gamma)$
3. Вычисление $r_{i,m} = -[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}]_{F(x)=F_{m-1}(x)}$
4. Обучение регрессии r_{jm} в цикле j = 1...m
5. В цикле j = 1...m вычисление $\gamma_{jm} = argmin\Sigma L(y_i, F_{m-1}(x_i) + \gamma)$
6. $F_m(x) = F_{m-1}(x) + \nu\Sigma\gamma_m I(x \in R_{jm})$

6

Время практики